INTERCALATOR-MEDIATED ASSEMBLY OF NUCLEIC ACIDS

A Dissertation
Presented to
The Academic Faculty

by

Eric D. Horowitz

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2009

INTERCALATOR-MEDIATED ASSEMBLY OF NUCLEIC ACIDS

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Roger Wartell
School of Biology
Georgia Institute of Technology

Dr. Steve Harvey
School of Chemistry and Biochemistry and
School of Biology
Georgia Institute of Technology

Date Approved: February 11, 2009

This thesis is dedicated to my family and to my wife Mary.
Without you none of this would have been possible.

ACKNOWLEDGEMENTS

A great advisor is truly hard to find. That said, I sincerely thank my advisor Prof. Nicholas Hud for being an inspiration in the lab, an outstanding scientist, and the best advisor any student could ask for. As a student under Prof. Hud, I have become an ever-more curious scientist, always looking for further answers and never accepting a result at face value. Indeed, Prof. Hud has helped me become the best scientist that I can be.

It takes a certain type of person to be an encouraging and helpful labmate, sharing successes (and failures) as they come, and always being great companions. For all those things, I thank my labmates—Aaron Engelhart, Heather Bean, Özgül Persil Çetinkol, Swapan Jain, Tumpa Sarkar, Igor Vilfan, Ragan Buckley, Jim Collins, Cathy Santai, and Irena Mamajanov. I would also like to thank the undergraduate researchers who have worked with me—including Seth Lilavivat, Kaycee Quarles, and Ben Holladay.

The members of my committee have held me and my work to high standards, and have expected much from me. Our discussions have helped focus and clarify my work, and for that I am grateful. Thank you to Prof. Loren Williams, Prof. Steve Harvey, Prof. Roger Wartell, and Prof. David Sherrill. Dr. Les Gelbaum and Prof. Markus Germann, thank you for teaching and refining my knowledge of NMR.

There is no way that I can give enough thanks to my wife Mary and to my family for being there as my own personal support team to help me get through these years of study. My grandparents provided a source of constant inspiration, while my parents and the rest of my family provided perpetual encouragement. Their love always helped me through the tough times. I especially want to thank Mary for always being there to keep

me level headed and to help me keep things in perspective. No words can express how much I appreciate you always being there. I love you!

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

NOMENCLATURE

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
d	Deoxyribo-
r	Ribo-
H-bonds	Hydrogen Bonds
G	Guanine
C	Cytosine
A	Adenine
T	Thymine
U	Uracil
mRNA	Messenger RNA
tRNA	Transfer RNA
p	Phosphate-
AFM	Atomic Force Microscopy
2′,5′ RNA	2′,5′-Linked Ribonucleic Acid
dH_2O	Distilled Water
UV-vis	Ultraviolet-Visible Spectrophotometry
CD	Circular Dichroism
K	Equilibrium Constant
ΔG°	Standard Free Energy Change
ΔH°	Standard Enthalpy Change
ΔS°	Standard Entropy Change

T	Temperature
T_M	Melting Temperature
ΔT_M	Melting Temperature Change
ICD	Induced Circular Dichroism
bp	Base Pair
mdeg	Millidegrees
LNA	Locked Nucleic Acids
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide Gel Electrophoresis
D_2O	Deuterium Oxide
NOESY	Nuclear Overhauser Effect Spectroscopy
COSY	Correlation Spectroscopy
TOCSY	Total Correlation Spectroscopy
HETCOR	Heteronuclear Correalation Spectroscopy
NOE	Nuclear Overhauser Effect
MD	Molecular Dynamics
DFT	Density Functional Theory
VMD	Visual Molecular Dynamics
AMBER	Assisted Model Building with Energy Refinement
ppm	Parts Per Million
Hz	Hertz
RMSD	Root Mean Square Deviation
PDB	Protein Databank
ITC	Isothermal Titration Calorimetry
Pyr	Pyrimidine

Pur	Purine
HPLC	High Pressure Liquid Chromatography
TEAA	Triethylammonium acetate
TEA	Triethylamine
MES	Morphylinoethyl Sulfonic Acid
ImCN	N-cyanoimidazole
CIAP	Calf Intestine Alkaline Phosphatase
aza3	Azacyanine 3
pRNA	Pyranosyl-RNA
TNA	Threose Nucleic Acid
GNA	Glycerol Nucleic Acid
FNA	Flexible Nucleic Acid
p**s**	Phosphorothioate
l	Locked Nucleic Acid Residue
PCR	Polymerase Chain Reaction

SUMMARY

The RNA World hypothesis suggests that RNA, or a proto-RNA, existed in an early form of life that had not yet developed the ability to synthesize protein enzymes. This hypothesis, by some interpretations, implies that nucleic acid polymers were the first polymers of life, and must have therefore spontaneously formed from simple molecular building blocks in the "prebiotic soup." Although prebiotic chemists have searched for decades for a process by which RNA can be made from plausible prebiotic reactions, numerous problems persist that stand in the way of a chemically-sound model for the spontaneous generation of an RNA World (*e.g.,* strand-cyclization, heterogeneous backbones, non-selective ligation of activated nucleotides). The Molecular Midwife hypothesis, proposed by Hud and Anet in 2000, provides a possible solution to several problems associated with the assembly of the first nucleic acids. In this hypothesis, nucleic acid base pairs are assembled by small, planar molecules that resemble molecules which are known today to intercalate the base pairs of nucleic acid duplexes. Thus, the validity and merits of the Molecular Midwife hypothesis can be, to some extent, explored by studying the effects of intercalation on the non-covalent assembly of nucleic acids.

In this thesis, I explore the role of the sugar-phosphate backbone in dictating the structure and thermodynamics of nucleic acid intercalation by using 2′,5′-linked RNA intercalation as a model system of non-natural nucleic acid intercalation. The solution structure of an intercalator-bound 2′,5′ RNA duplex reveals structural and thermodynamic aspects of intercalation that provide insight into the origin of the nearest-neighbor exclusion principle, a principle that is uniformly obeyed upon the intercalation

of natural (i.e. 3′,5′-linked) RNA and DNA. I also demonstrate the ability of intercalator-mediated assembly to circumvent the strand-cyclization problem, a problem that otherwise greatly limits the polymerization of short oligonucleotides into long polymers. Together, the data presented in this thesis illustrate the important role that the nucleic acid backbone plays in governing the thermodynamics of intercalation, and provide support for the proposed role of intercalator-mediated assembly in the prebiotic formation of nucleic acids.

CHAPTER 1

INTRODUCTION

1.1. THE ORIGIN OF LIFE AND THE RNA WORLD HYPOTHESIS

The so-called "central dogma" of molecular biology, which was first proposed by Crick [1], is the concept that deoxyribonucleic acid (DNA) stores information for the genetic traits of living cells. For protein synthesis, ribonucleic acid (RNA) is created (transcribed) from a DNA template, which then encodes for proteins (via mRNA). The utility of transfer RNA (tRNA) and the multi-molecular protein/RNA complex, known as the ribosome, carryout the translation of mRNA codes into proteins [2]. Each step of the above scenario is highly regulated by the incorporation of numerous protein enzymes and transcription factors, as well as by the DNA and RNA structures themselves.

The realization that proteins and nucleic acids represent the informational macromolecules of living organisms quickly resulted in a 'chicken and the egg' dilemma. That is, which macromolecule came first: proteins, which are needed to act as the cells' catalysts, or nucleic acids, which encode for the proteins themselves? In 1953, Stanley Miller demonstrated the formation of amino acids, the building blocks of proteins, by simply applying an electric discharge to a mixture of prebiotic gases (NH_3, CH_4, H_2O, and H_2) [3]. John Oró subsequently showed that adenine, one of the nucleic acid bases, could be created from the heating of hydrogen cyanide [4]. This experimental evidence, as well as others of similar merit [5], suggested that the prebiotic synthesis of both proteins and nucleic acids is possible, which further emphasized the dilemma regarding which came first.

Considerable support for the primacy of nucleic acids over proteins was provided in the early 1980s by the observation that some RNA molecules can catalyze reactions [6,7]. Thus, RNA can both store genetic information and control chemical reactions. This realization led to the so-called 'RNA-World' hypothesis: which states that RNA could have acted as both the information carrier and molecular machinery required to commence life [8]. Since the time that the first natural RNA enzymes (or ribozymes) were discovered, numerous natural and synthetic ribozymes have been identified that are capable of performing a myriad of catalytic functions [9-12].

In addition to the investigation of ribozymes that might have facilitated the origin of life, considerable research has also been dedicated toward understanding how the first RNA polymers could have been formed from simple precursors without the use of protein enzymes. It stands to reason; the natural selection of RNA sequences that could have started life would have required a starting "pool" of RNA polymers. Orgel and coworkers demonstrated that monomeric G nucleotides could be assembled on a poly(C) template [13]. In order to form the high energy phosphodiester bond, the G nucleotides required activation, by converting the 5′ phosphate into a 5′ phosphor-2-methylimidazolide, for example. Although it was of great importance to demonstrate that activated nucleotides can polymerize along a template strand without the aid of a protein enzyme, Orgel and coworkers observed that only certain conditions produced polymers. For example, the addition of other bases into the template greatly reduced product formation. Furthermore, with regards to models for the origin of RNA, the work of Orgel and co-workers always required a pre-existing template.

More recently, Ferris and coworkers have shown that mineral catalysts, such as the clay montmorillonite, can provide a surface for the assembly of methylimidazole-activated monomers without a template strand [14]. While some researcher consider this result a the solution to the problem of generating long nucleic acid polymers in a prebiotic world, surface-directed polymerization of mononucleotides still has limitations that leave open the question of whether or not mineral surfaces were involved in proto-RNA polymerization. For example, the nature of the electrostatic interactions that the nucleic acid monomers have with the surface implies that the products of the reaction may become increasingly difficult to remove from the surface as they increase in length. Additionally, the reaction is not specific to nucleic acids capable of further replication [15]. Therefore, another problem with forming the first RNA polymers is not only being able to build polymers from prebiotic starting materials, but also replication of these polymers. These requirements illustrate some of the questions that remain regarding the origin of the first RNA polymers. How can nucleotides assemble in a fashion that will include only those that are capable of further propagating genetic information? Considerations of contemporary nucleic acid structure may provide clues to the process that originally gave rise to RNA, or its predecessor.

1.2. NUCLEIC ACID STRUCTURE

The structure of nucleic acids has been studied for over 50 years [16-20]. Nucleic acids are polymers that are composed of monomeric units, called nucleotides [21]. A nucleotide is composed of a ribose sugar attached to a hetero-aromatic base. Also attached to the sugar is a phosphate group that connects monomer nucleotides through

phosphodiester linkages to form polymers. In natural DNA and RNA, these linkages are 3′,5′ (*i.e.,* the phosphate group is connected between the 3′ position of one nucleotide and the 5′ position of the next along the polymer backbone). Through base pairing nucleic acids can then form the familiar double helix, in which base paired strands are orientated anti-parallel to each other.

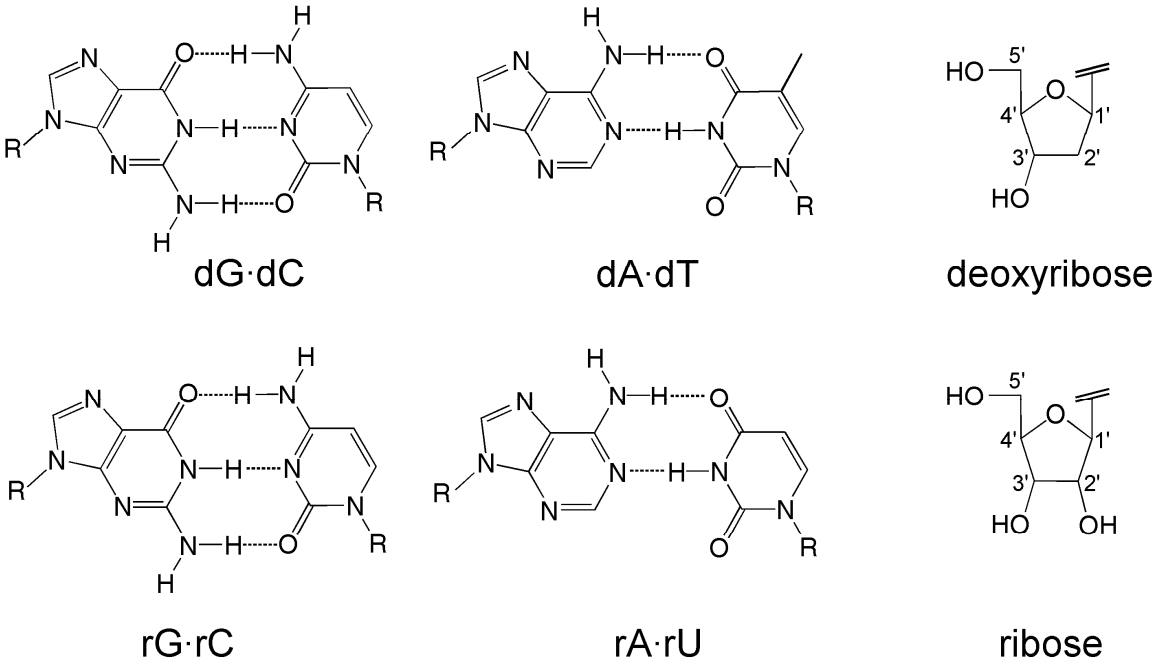


Figure 1.1. The chemical structures of the Watson–Crick base pairs as they appear in DNA and RNA (denoted d and r respectively).

There are several interactions that stabilize a nucleic acid double helix. First, there are hydrogen bonds (H-bonds) between the nitrogenous bases. These bases are either purines or pyrimidines. In Watson-Crick base pairs, the pyrimidines thymine and uracil form two H-bonds with the purine adenine, while the pyrimidine cytosine can form three H-bonds with the purine guanine (Figure 1.1). In retrospect, the work of Chargaff *et al.* first suggested that there must be a physical pairing between the pyrimidine and

purine bases, as their data revealed universal ratios between the composition of purine and pyrimidine bases for a wide range of organisms [17]. The combined efforts of Watson, Crick, Franklin, and Wilkins provided a model for the three dimensional structure of DNA based upon X-ray fiber diffraction data, which was the first proposal of the double helical structure [19,22,23]. Two decades later, the Watson–Crick model of the DNA double helix was confirmed [24] when the first high-resolution single-crystal X-ray structure was determined for a DNA duplex [25].

In addition to base pairing, DNA and RNA duplexes are held together by inter- and intrastrand stacking interactions [21]. The hydrophobic base pairs stack in aqueous solution to exclude water (*i.e.,* the hydrophobic effect). Additionally these interactions are stabilized by favorable π-π interactions. Lastly there are intimate water and metal ion contacts along the edges of the nucleic acid bases in the major and minor grooves that also participate in favorable electrostatic interactions with the negatively charged phosphate groups [26,27]. While all of these interactions are necessary for duplex formation, π-π interactions have been observed *in silico* to be critical to duplex formation [28].

The DNA double helix is typically in the B-form (Figure 1.2) [21]. B-form helices have their base pairs stacked perpendicularly to the helical axis, with a helical rise of 3.4 Å. Other structural features that characterize the B-form helix are the 36º helical twist and minimal displacement of the base pairs from the helical axis [21]. Additionally, the deoxyribose sugars in DNA are primarily in the C2′ *endo* sugar pucker conformation with an inter-phosphate distance of *ca.* 7 Å. The B-form helix has two distinct grooves (Figure

1.2). The larger of these two grooves is termed the major groove and the smaller is termed the minor groove.

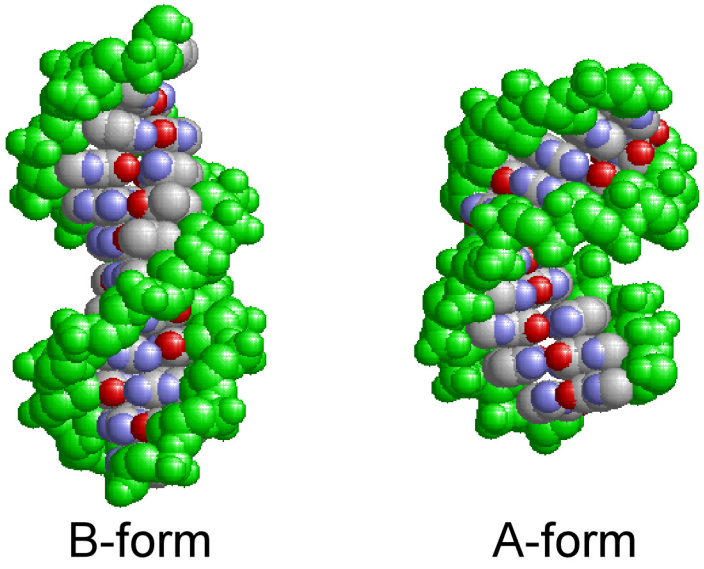


Figure 1.2. Sample B-form [29] and A-form DNA [30] structures. The sugar-phosphate backbone is shown in green while the base pairs are stacked up the center of the helices.

RNA is composed of very similar constituent parts, but the helical structure is considerably different from DNA [21]. The main difference between DNA and RNA is at the nucleotide structure level. RNA has a ribose sugar whereas DNA has a 2′-deoxyribose sugar. This seemingly minor structural difference causes the RNA double helix to be in the A-form helical conformation (Figure 1.2). A-form helices are slightly more wound than B-form helices, tilting the bases beyond perpendicular from the helical axis, and decreasing the helical rise to *ca.* 3 Å [21]. Additionally, the base pairs are displaced from the center of the helical axis by *ca.* 4 Å, creating a hole down the center of the helix. Nucleic acids in the A-form tend to have C3′ *endo* sugar puckers [21]. While DNA is

typically thought of as being primarily in the B-form, it was shown by Franklin that, under conditions of dehydration, DNA can also be in the A-form (Figure 1.2) [22].

1.3. NUCLEIC ACID-LIGAND BINDING INTERACTIONS

Duplex nucleic acids can be thought of as having two distinct regions in which small molecules can bind. First, small molecules can bind to the hydrophobic core that is formed along the helical axis by stacked nucleic acid base pairs. Second, small molecules can bind to the major and minor grooves as they are hydrophilic crevices that are rich in sequence specific information. For example, in the major groove the base pairs present patterns of H-bond donors and acceptors which are then recognized by DNA-binding proteins [2]. Structural studies of small molecules bound to nucleic acids [31,32] demonstrate that generally the observed association constant can not be attributed to the interaction of a small molecule with solely one region of the nucleic acid duplex. Given the different substituents associated with the various nucleic acid-binding small molecules, often small molecules will interact with both the duplex hydrophobic core as well as the grooves [33].

As mentioned above, the major and minor grooves of DNA and RNA are rich in H-bond donor and acceptor sites. Wartell and Wells first proposed that netropsin, a polyamide with H-bond donors and acceptors along its chain, binds to DNA in one of the grooves [34]. The base-specific interactions between netropsin and the minor groove of DNA inspired Dervan and coworkers to develop a series of polyamides that can be individually designed to bind to specific DNA sequences [35-37].

Small molecules that insert planar groups between the base pairs of nucleic acids are termed intercalators. The backbone of DNA and RNA is dynamic and flexible [38,39], which allows two adjacent base pairs to destack and make room for such a planar aromatic system to insert between the two base pairs (Figure 1.3) [40]. Lerman first proposed the intercalation binding mode when he observed that acridines associate with DNA in a mode perpendicular to the helical axis [41]. X-ray diffraction of DNA co-crystallized with acridines confirmed that intercalators extend the DNA helix 3.4 Å enough to fit a new π-system between two base pairs [42].

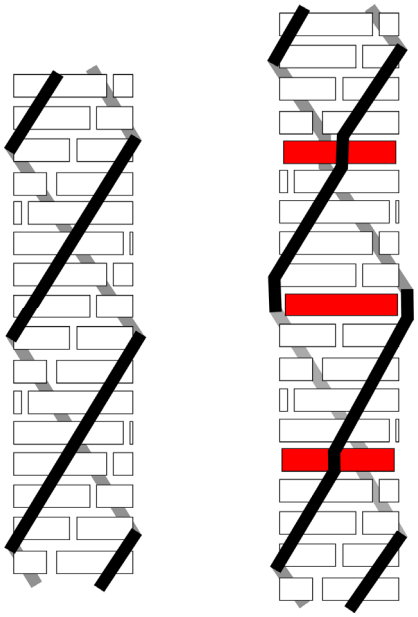


Figure 1.3. Cartoon representation of intercalation of a small molecule (red) into a native duplex (left). This binding mode is associated with an extending of the helix at the site of intercalation (right).

Many small molecule intercalators such as idarubicin [43] and actinomycin D [44] are used clinically. Thus, an in-depth understanding of the binding of these molecules to nucleic acids could facilitate the rational design of new pharmaceutical drugs. Furthermore, in spite of forty years of research there are still several features of

intercalation that are not yet not fully understood. For example, intercalators can bind at a maximum of one small molecule to every two base pairs [45]. This is known as the nearest-neighbor exclusion principle, but the underlying physical basis for this principle remains under debate [46-52].

1.4. INTERCALATORS AND THE ORIGIN OF RNA POLYMERS

As mentioned in Section 1.1, the assembly mechanism for the first RNA-like molecules is still an open question. In 2000, Hud and Anet proposed a solution to some of the problems associated with the abiotic formation of RNA-like polymers [53]. These authors proposed that a small molecule template (or 'Molecular Midwife') that is similar in size and shape to the Watson–Crick base pairs could have provided a hydrophobic surface for Watson–Crick base pair assembly in solution. Once the bases are assembled in an alternating stack of midwife molecules and base pairs, the backbone could assemble along the edges (Figure 1.4). This hypothesis is rooted in the fundamental forces that hold nucleic acids together, such as π-stacking and H-bonding between base pairs, rather than merely the chemical functionality of activated nucleotides. In addition, simple drying (*e.g.,* pond evaporation in the sun) could have facilitated the assembly of the bases with midwives and the formation of the backbone, as many of the bonds formed in nucleic acids are formed through dehydration reactions (*e.g.,* phosphodiester, glycosidic). Similarly, a cool evening dew could rehydrate the nucleic acids and dilute them beyond the dissociation constant of the molecular midwife could then release the small molecules.

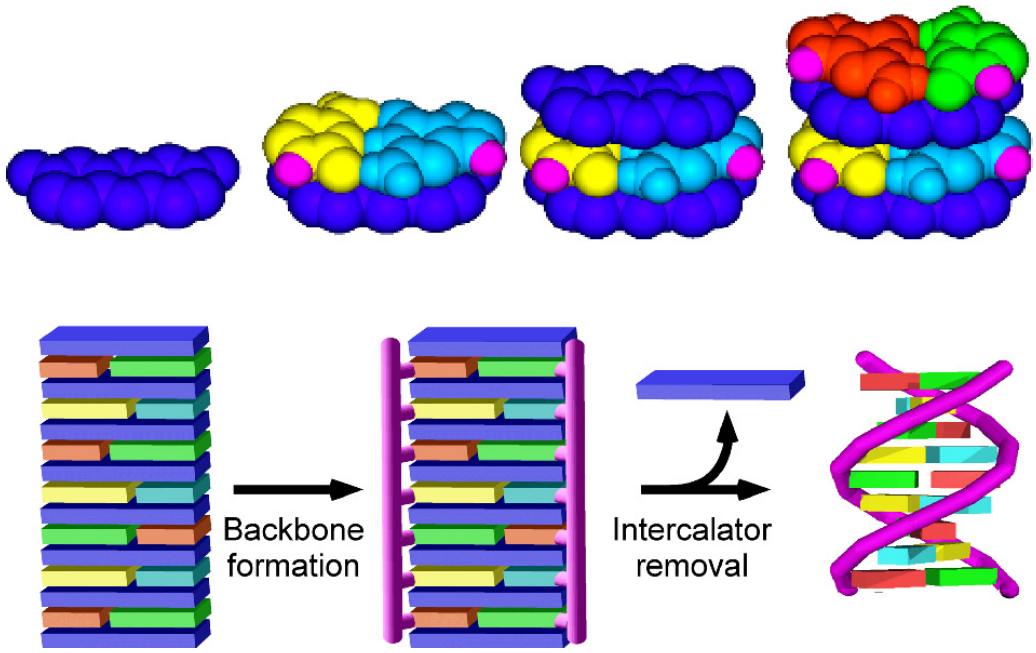


Figure 1.4. Summary of the molecular midwife hypothesis [53,54]. Intercalator-like small molecules (or 'Molecular Midwives') act as a hydrophobic template for the non-covalent association of the nucleic acid base pairs (top). Backbone formation along the edges of the midwife-base pair stack is driven by dehydration (bottom). Rehydration then allows the release of the midwife, creating a native nucleic acid duplex.

DNA-intercalating ligands, such as proflavine, represent reasonable approximations to the size, shape, and molecular properties of the molecular midwife molecules proposed by Hud and Anet. While the self-assembly of the free nucleic acid bases, as shown in Figure 1.4, has not yet been demonstrated, the group of intercalators studied to date are by no means exhaustive. The association of mononucleotide base pairs in aqueous solution, somewhat akin to the assembly proposed by Hud and Anet, was recently observed by Sawada *et al.* in the presence of a hydrophobic clathrate cage [55], albeit by anti-Hoogsteen base pairing [56]. During the past several years, the Hud laboratory has provided significant evidence of the utility of nucleic acid intercalation in the assembly of nucleic acid oligonucleotides. For example, Polak and Hud showed that

the addition of coralyne, a crescent shaped intercalator that binds specifically to triplex structures [57], disproportionates poly(dA)·poly(dT) duplexes into the triplex poly(dT)·poly(dA)·poly(dT) and a non-canonical poly(dA)·poly(dA) structure [58]. It was also observed that the oligonucleotides $d(A_{16})$ and $d(T_{16})$ follow the same trend in the presence of coralyne [59]. In addition, Persil *et al*. observed that $d(A_8)$, $d(A_{16})$, and $d(A_{32})$ assemble into anti-parallel non-covalent assemblies only observed in the presence of coralyne [60].

In 2004 Jain *et al*. demonstrated that $d(T_3)$ with a 3′ phosphorothioate ligates to $d(T_4)$ with a 5′ iodo leaving group on a $d(A_{16})$ template only in the presence of the intercalator proflavine [61]. At low micromolar concentrations the association of $d(T_3)$ and $d(T_4)$ on the $d(A_7)$ template is small without the free energy contribution from the bound intercalator even at 4°C. When the intercalator is present, a 1000-fold increase in ligation was observed because of the added duplex stability. Clearly, intercalation of nucleic acids could have been very useful in the assembly and formation of the first RNA polymers.

As stated above, the molecular midwife hypothesis takes advantage of many features that are common to nucleic acids as they are today. Forming the nucleic acid phosphodiester bond is no exception to this trend. However, phosphodiester bond formation requires a large energetic input. This means that in the template directed synthesis studies discussed above, the phosphate is typically chemically activated [14,61-65]. An alternate backbone may be utilized in the origin of life that is thermodynamically driven rather than kinetically driven. Bean *et al*. showed that glyoxylic acid bridges the 5′ and 3′ hydroxyls of two nucleotides by an acetal linkage [66]. The bond formed using

glyoxylic acid is the same length as the phosphodiester linkage, has a negative charge, and is thermodynamically controlled. This example, as well as others, provides support for the hypothesis that before the RNA world there must have been a pre-RNA world, which used a different RNA-like backbone and perhaps even different base pairs [67-69].

CHAPTER 2

THERMODYNAMICS OF LIGAND-2′,5′ RNA ASSOCIATION[1]

2.1. INTRODUCTION

Small molecule intercalation of DNA and RNA has been the subject of biophysical and biochemical investigations for over forty years [41,70-72]. Despite these studies, and despite the recognized importance of nucleic acid intercalation in medicine and molecular biology, fundamental questions persist regarding the energetics of intercalation [38]. The dramatic unwinding and extending of DNA and RNA duplexes upon intercalation immediately suggests that the nucleic acid backbone should play a significant role in dictating the free energy of intercalation. However, the free energy contribution of the backbone is difficult to appreciate given the intertwined energetics associated with intercalation (*e.g.,* hydrophobic base stacking and solvent effects) [38]. Furthermore, previous intercalation studies are almost exclusively limited to natural DNA and RNA, making it difficult to predict how changes in backbone structure will promote or hinder intercalation. The data presented here, which was also published in 2006 [73] includes binding studies of known intercalators to a duplex of 2′,5′-linked RNA, a nucleic acid that is arguably the closest chemical analog of standard RNA, but with distinct backbone atom connectivity (Figure 2.1).

---

[1] The majority of the work in this chapter was published previously (Horowitz, E.D. and Hud, N.V. (2006) "Ethidium and proflavine binding to a 2′,5′-linked RNA duplex." *J. Am. Chem. Soc.*, **128**" 15380-15381.)

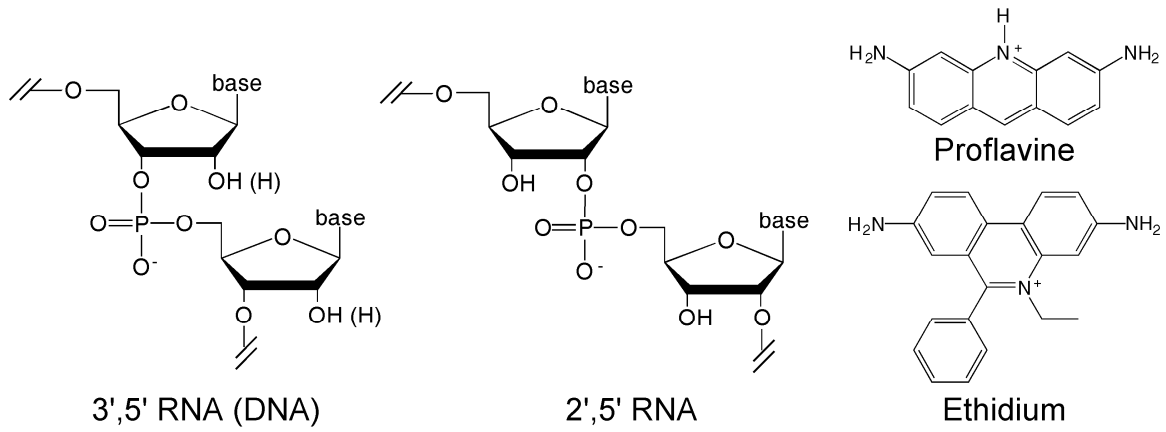


Figure 2.1. Structures of 3′,5′ RNA, 3′,5′ DNA, and 2′,5′ RNA as well as the Watson–Crick base pair intercalators proflavine and ethidium.

Proflavine and ethidium (Figure 2.1) were chosen for this initial study of 2′,5′-linked RNA binding because their respective interactions with natural DNA and RNA has been studied in detail. Additionally, proflavine intercalates in nucleic acids with limited contacts beyond base-intercalator stacking, whereas the pendant ring of ethidium protrudes into the minor groove making extensive van der Waals interactions in the minor groove[32,74].

2.2. EXPERIMENTAL PROCEDURES

2.2.1. Materials

3′,5′ DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and were used as received. 3′,5′ RNA oligos were purchased from Dharmacon Inc. (Lafayette, CO) and were used as received. 2′,5′ RNA oligos were synthesized in house on an automated Expedite synthesizer using standard phosphoramidte chemistry and the corresponding 3′-TBDMS-protected 2′-

phosphoramidites (ChemGenes). Following deprotection, oligonucleotides were separated by length on a 20% denaturing polyacrylamide gel. The band corresponding to the full length product was cut from the gel, extracted from the gel matrix by the crush-and-soak method, loaded onto an anion exchange column, washed with 150 mM NaCl, and eluted with 2.5 M NaCl. The oligonucleotides were then ethanol precipitated, dried, resuspended in dH_2O and passed over a 1 m Sephadex G-10 column. Combined fractions containing the purified product (sodium salt) were pooled, lyophilized, and resuspended in dH_2O.

Extinction coefficients for 3′,5′ RNA and 2′,5′ RNA oligonucleotides were determined by cleavage in 0.3 M NaOH [75] and comparison to nucleotide monophosphate extinction coefficients [76]. Extinction coefficient determination was reproducible with less than 8% variation. Extinction coefficients for DNA oligonucleotides were provided by the manufacturer.

Table 2.1. Nucleic acid extinction coefficients

Type	**Sequence**	**ε_{260} (M^{-1} cm^{-1})**
3′,5′ RNA	5'-CCG GCC GCG CGC	67,200 ± 8%
3′,5′ RNA	5'-GCG CGC GGC CGG	76,800 ± 6%
2′,5′ RNA	5'-CCG GCC GCG CGC	63,200 ± 6%
2′,5′ RNA	5'-GCG CGC GGC CGG	74,400 ± 6%
3′,5′ DNA	5'-CCG GCC GCG CGC	98,700
3′,5′ DNA	5'-GCG CGC GGC CGG	105,700

Ethidium bromide (Fisher) and proflavine hemisulfate (Sigma) were used as received and dissolved in dH_2O. Extinction coefficients used to determine stock solution concentrations were ε_{480} = 5 800 M^{-1} cm^{-1} for ethidium and ε_{444} = 41 000 M^{-1} cm^{-1} for proflavine.

2.2.2. Fluorescence Data Collection and Analysis

All fluorescence measurements were performed on a Shimadzu RF-5301PC spectrofluorophotometer at 25ºC. Sample solution conditions: 1× BPE buffer (2 mM NaH_2PO_4, 6 mM Na_2HPO_4, 1 mM Na_2EDTA) pH 7, 100 mM NaCl, unless stated otherwise. Titrations were carried out by making incremental additions of a nucleic acid stock solution to a larger volume solution containing 1 μM ethidium or proflavine. The stock nucleic acid solutions also contained the respective small molecule at 1 μM, in order to avoid dilution of the small molecule over the course of the titration. Ethidium was excited at 510 nm with a 5 nm excitation bandwidth and detected with a 5 nm emission bandwidth. Proflavine fluorescence was excited at 455 nm with a 1 nm excitation bandwidth and detected with a 3 nm emission bandwidth.

Equilibrium constants were derived from fluorescence titration data as described previously by Qu and Chaires [77]. Briefly, integrated florescence intensity measurements were fit by the method of least squares using the equations:

$$KC_b^2 - C_b(KS_0 + KD_0 + 1) + KS_0D_0 = 0$$

$$F = F^0(C_t - C_b) + F^bC_b$$

where K is the association constant in (M^{-1}), C_b is the concentration of bound ligand, S_0 is the total binding site concentration, C_t is the total concentration of small molecule, D_0 is the total ligand concentration, F is observed fluorescence at each titration point, F^0 is fluorescence intensity of the free small molecule, and F^b is the fluorescence of the bound species.

2.2.3. Circular Dichroism and UV-vis Spectrophotometry

UV-vis spectra were acquired on a Hewlett Packard 8453 diode array spectrophotometer equipped with an Agilent 89090A Peltier temperature controller. UV melting profiles were produced by monitoring absorbance at 260 nm while increasing the temperature at a rate of 1ºC min^{-1} from 5 to 95ºC. CD spectra were acquired on a JASCO J-810 CD spectropolarimeter at 25ºC. All CD spectra were baseline subtracted with a separately acquired buffer spectrum.

2.3. RESULTS AND DISCUSSION

2.3.1. Proflavine and Ethidium Association with 2′,5′ RNA, DNA, and RNA

The changes in fluorescence intensities of proflavine and ethidium upon base pair intercalation, can be used to determine binding constants [77]. Fluorescence titration studies reveal that proflavine binds a 2′,5′ RNA duplex with an association constant of $K = 3.9 \times 10^5$ M^{-1}. This is more favorable than that exhibited by DNA and RNA duplexes of the same length and sequence (Figure 2.2) [73]. In contrast, ethidium binds the same 2′,5′ RNA duplex with $K = 0.4 \times 10^5$ M^{-1}, which is *less* favorable than that exhibited by DNA and RNA (Figure 2.2) [73]. *K* values measured for RNA under identical conditions reveal that changing the RNA backbone linkage from 3′,5′ to 2′,5′ results in a 25-fold increase in the *K* for proflavine binding, corresponding to a ΔΔG° of -2 kcal mol^{-1} at 25ºC. This is an appreciable increase, considering that the previously reported ΔG° for proflavine binding to polymer RNA has been reported as *ca.* -8 kcal mol^{-1} at 25ºC [78]. The same backbone change results in a 2-fold decrease in the *K* for ethidium binding, corresponding to a ΔΔG° of +0.4 kcal mol^{-1} at 25ºC.

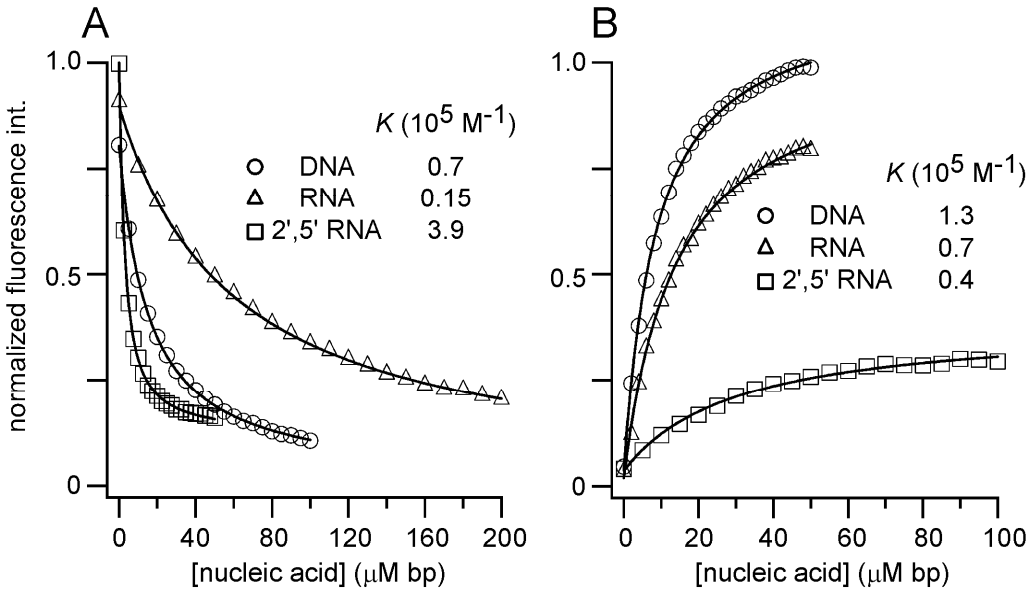


Figure 2.2. (A) Fluorescence intensity measurements for 1 μM solutions of proflavine in the presence of increasing concentrations of DNA, RNA and 2′,5′-linked RNA duplexes with the nucleotide sequence 5′-CCGGCCGCGCGC and its complement. (B) Same as panel A, except for 1 μM solutions of ethidium. *K* values were derived from the curve fits shown. Samples were 25°C, 100 mM NaCl, 1× BPE, pH 7 [73].

It was shown previously that the association of intercalators with DNA and RNA are dependent on the salt concentration [38,51,79]. This effect has been attributed to the extending transition observed in intercalation. Separation of the negatively charged phosphates decreases the axial charge density in the vicinity of an intercalation site. It is because of this structural transition that counter-ions are released [79]. Therefore, an increase in salt concentration decreases the association constant of intercalators. In addition, many intercalators are positively charged at neutral pH. Positively charged ligands can also interact with the backbone phosphates, releasing counterions [33,38].

Proflavine and ethidium association with 2′,5′ RNA responds very similarly to salt concentration as 3′,5′ DNA and RNA (Figure 2.3) [73]. Record and Manning's theory [80,81] describes the association of positively charged ligands associating with

polyanionic molecules such as a large strand of DNA. Using this theory, the free energy contribution to the total association free energy can be calculated. Although the same approximations made by Record and Manning cannot be applied to short duplexes, the slopes of the lines in Figure 2.3 are similar to what is observed for monocationic intercalators binding to polymeric DNA [79].

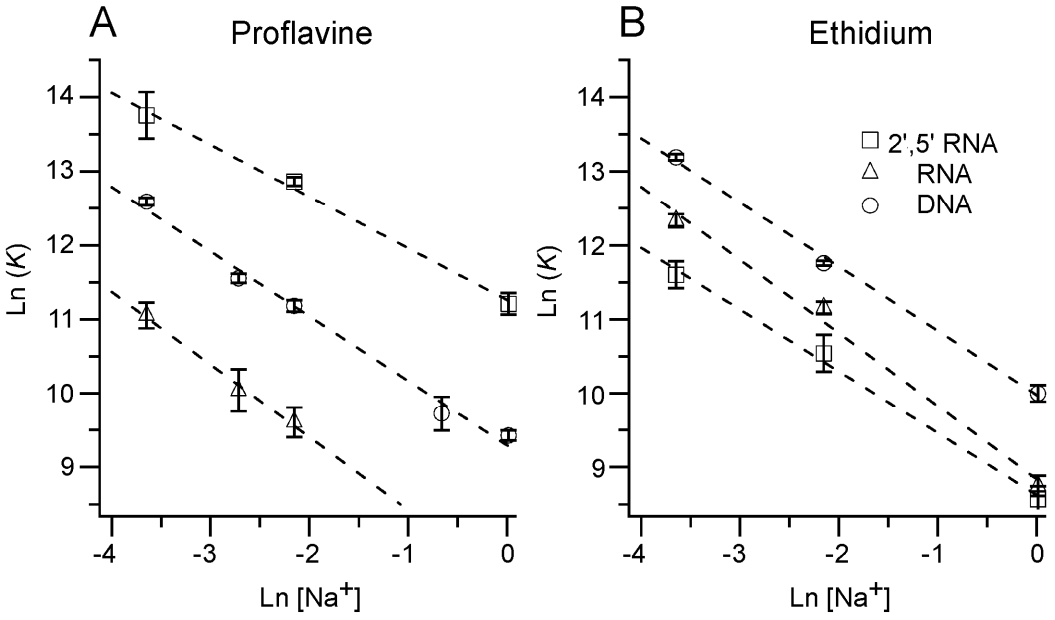


Figure 2.3. A) Plots of the effect of increasing sodium concentration on the association of proflavine with 2′,5′ RNA, RNA, and DNA duplexes with the sequence 5′-CCGGCCGCGCGC. Dashed lines represent best fit lines and error bars represent the standard deviation from three titrations. B) The same as panel A only for ethidium. Samples were 25°C, 1× BPE, pH 7, with the indicated salt concentration [73].

2.3.2. Proflavine- and Ethidium-Bound 2′,5′ RNA have Spectral Properties Similar to Natural Nucleic Acids

The association of ligands such as ethidium and proflavine with DNA changes the ligand maximum absorbance wavelength in the visible spectrum. In general intercalators have a red shift in their visible absorbance band (479 nm and 444 nm for ethidium and

proflavine respectively) and a decrease in the intensity (hypochromism) [70,82,83]. These changes are caused by transfer of the ligand chromophore from the aqueous solvent into the relatively hydrophobic environment between the base pairs. This same trend is observed for the association of ethidium and proflavine with 2′,5′ RNA (Figure 2.4). The 444 nm band of free proflavine exhibits a 22 nm red-shift and 8% hypochromism upon binding to 2′,5′ RNA; changes that are compatible to those exhibited by proflavine upon binding to DNA (15 nm, 17%) and RNA (13 nm, 18%) [73]. The 479 nm band of ethidium exhibits a 15 nm red-shift upon binding to 2′,5′ RNA (compared to 38 and 31 nm for DNA and RNA, respectively) and 17% hypochromism (compared to 21% and 24% respectively) [73].

The mixture of proflavine with 2′,5′ RNA produces an increase in the melting temperature as does ethidium (Figure 2.5). The similarities between the data presented for the three nucleic acid duplexes are suggestive of a similar binding mode. Ethidium binding increases the 2′,5′ RNA duplex melting temperature (T_M) by 6°C, from a free duplex T_M of 42°C (compared to 3°C and <1°C for DNA and RNA, respectively). On the other hand, proflavine binding increases the T_M by 13°C (compared to 2°C and 1°C for DNA and RNA, respectively). Although the ΔT_M observed for proflavine binding to 2′,5′ RNA is striking compared to the T_M values observed for DNA and RNA, the free duplex T_M for 2′,5′ RNA is lower than that of DNA (73°C) and RNA (86°C). Therefore, it is difficult to make a direct comparison of the ΔT_M.

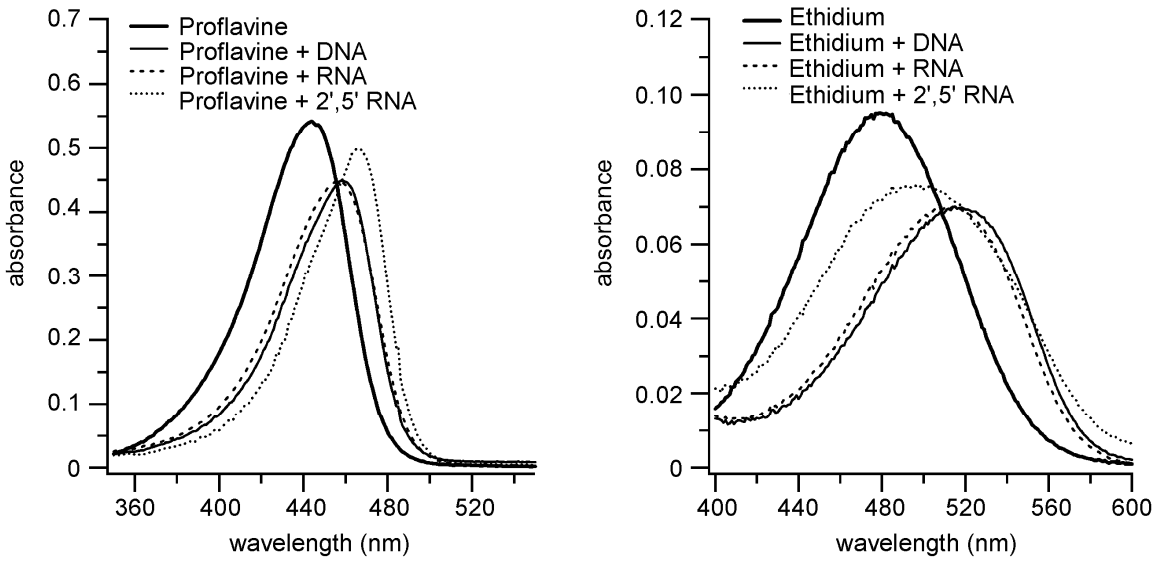


Figure 2.4. Visible absorbance spectra of solutions containing proflavine and ethidium (16 μM) and 2′,5′ RNA, RNA, and DNA (160 μM in bp) [73].

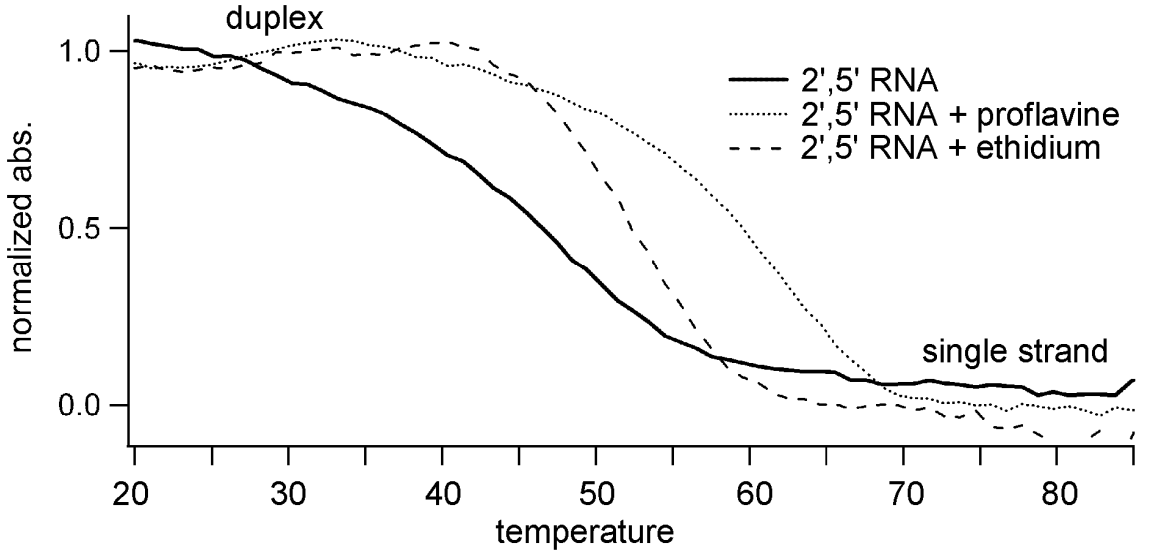


Figure 2.5. Melting profiles for 2′,5′-linked RNA with and without proflavine and ethidium, respectively. Solutions were 40 μM nucleic acid in bp, 20 μM in small molecule (where applicable), 1× BPE, 100 mM NaCl, pH 7. All experiments used the same dodecanucleotide sequence 5′-(CCGGCCGCGCGC)-2′ and its complement. Melting profiles were generated from 260 nm absorbance and normalized to pre- and post-transition baselines [73].

Circular dichroism is similar to UV-vis in that it is able to detect changes in the electronic properties of the nucleic acid bases. Whereas UV-vis detects changes in absorbance, CD detects changes in the rotation of circularly polarized light. CD spectra of nucleic acids appear as a series of positive and negative peaks centered about the nucleic acid UV absorbance at *ca.* 260 nm. Achiral ligands such as proflavine do not produce a CD spectrum. On the other hand, when bound to a chiral molecule such as DNA, proflavine produces a set of induced CD (ICD) bands in the UV region as well as the visible region of the spectrum [84]. Similar to natural DNA and RNA, 2′,5′ RNA produces a change in the UV region of the CD spectrum as well as an ICD band in the visible region when bound to proflavine (Figure 2.6) [73]. The visible ICD bands observed on the binding of proflavine to 2′,5′ RNA are very similar to that observed for DNA and RNA (albeit *ca.* 6-fold larger than the RNA ICD), implying a similar binding mode [73]. Ethidium produces very little change in the UV region of the CD spectrum and no change in the visible region for all three nucleic acid duplexes (Figure 2.7). This is not entirely surprising considering the 10-fold smaller molar extinction coefficient compared to proflavine in the visible region.

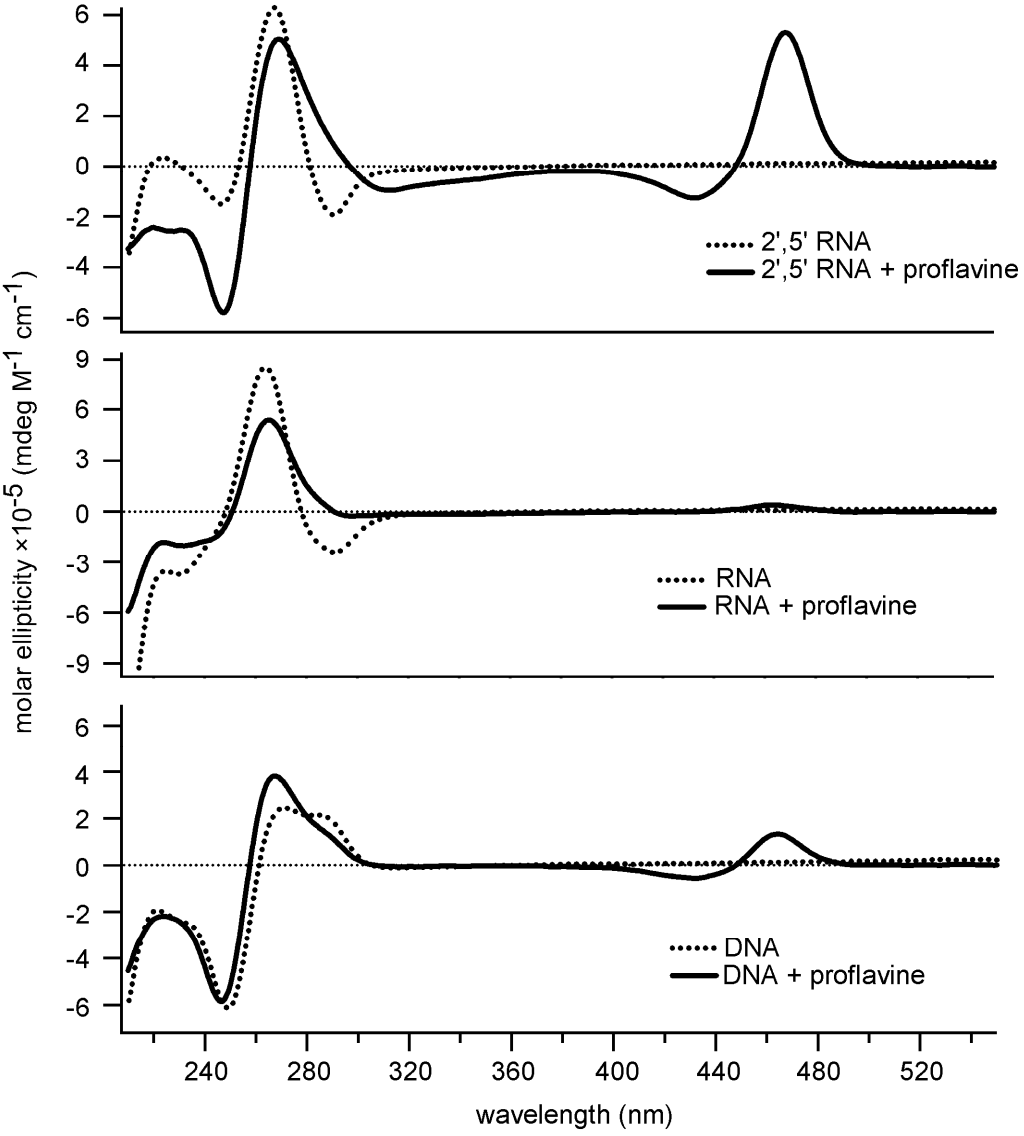


Figure 2.6. Circular Dichroism spectra for 2′,5′ RNA, 3′,5′ RNA, and 3′,5′ DNA of the sequence 5′-(CCGGCCGCGCGC)-3′ (or 2′) and its complement with and without proflavine. Concentrations were 80 μM in nucleic acid base pairs (bp) and 40 μM ethidium. Spectra were collected at 25°C in 1× BPE, 100 mM NaCl, pH 7.

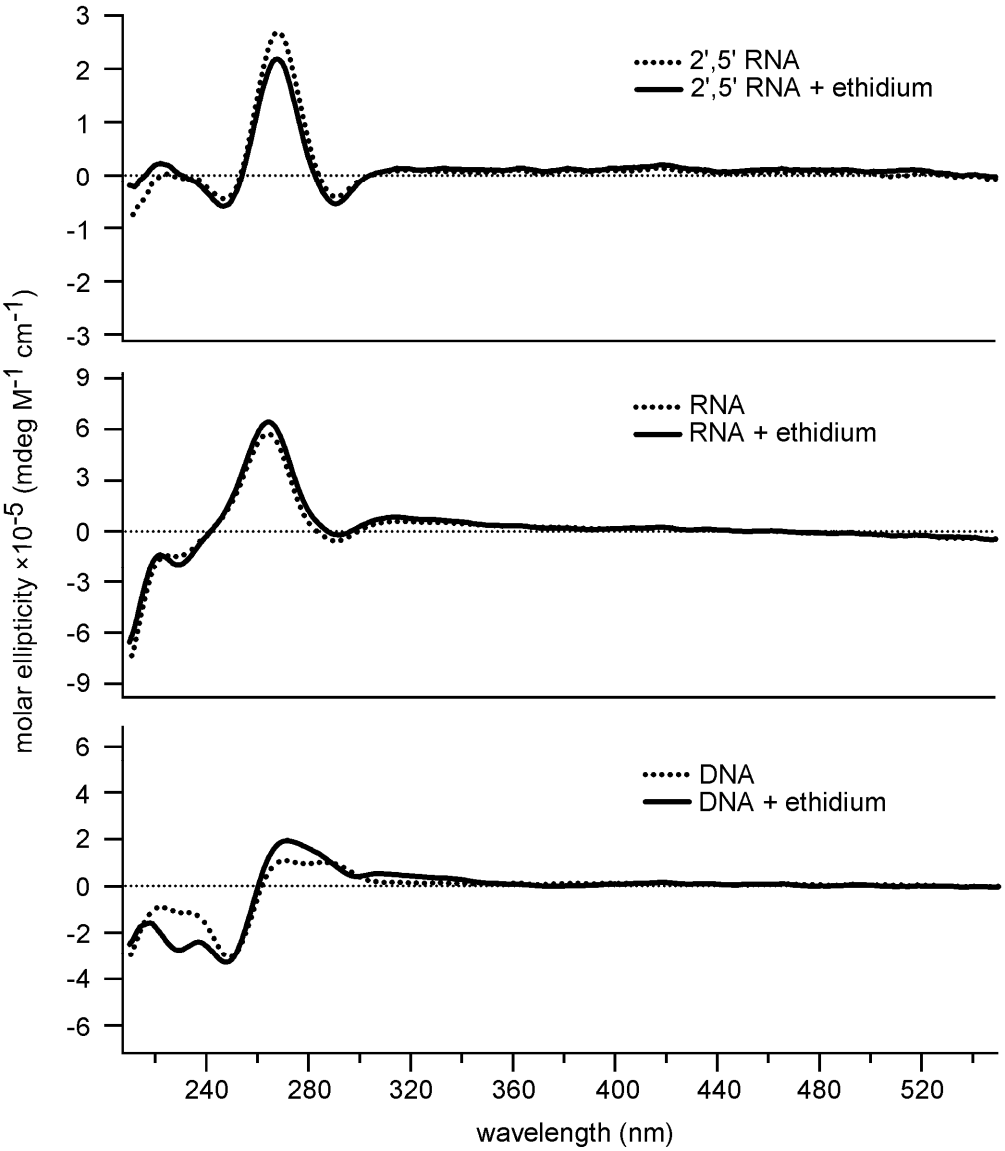


Figure 2.7. Circular Dichroism spectra for 2′,5′ RNA, 3′,5′ RNA, and 3′,5′ DNA of the sequence 5′-(CCGGCCGCGCGC)-3′ (or 2′) and its complement with and without ethidium. Concentrations were 40 μM in nucleic acid bp and 20 μM ethidium. Spectra were collected at 25°C in 1× BPE, 100 mM NaCl, pH 7.

2.3.3. The Association and Binding Stoichiometry of Proflavine with 2′,5′ RNA

Intercalators have been shown to bind to DNA at a maximum ratio of one small molecule to two base pairs (nearest-neighbor exclusion [45]) (Figure 2.8). One method of determining binding stoichiometry is by using Job's method of continuous variation [85]. In this method, the mole fraction of ligand and the mole fraction of nucleic acid are varied while maintaining a constant total number of moles (ligand + DNA) in the solution. An inflection point in the observed signal indicates the binding stoichiometry. For convenience of analysis, the number of moles of nucleic acid was defined in terms of nearest-neighbor binding sites (four bases or two base pairs). Therefore, the observation of an inflection point at R = 0.5 indicates a ratio consistent with the nearest-neighbor exclusion principle, where R = moles of proflavine / (moles of proflavine + moles of nucleic acid binding site).

The visible ICD band observed for the proflavine association with 2′,5′ RNA is ideal for a Job plot because of its relatively high intensity. Even at low R values a significant ICD band was observed. The Job plot in Figure 2.9 shows that there is an inflection point at *ca.* R = 0.5. Therefore, the stoichiometry of the association of proflavine and 2′,5′ RNA is consistent with the nearest-neighbor exclusion principle [73].

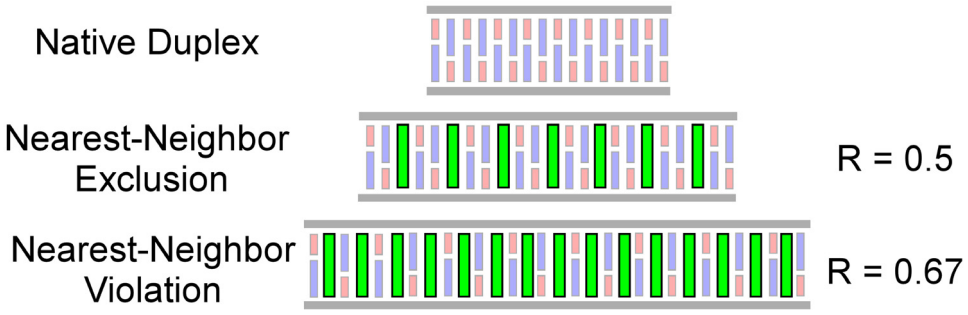


Figure 2.8. Schematic of the nearest-neighbor exclusion principle.

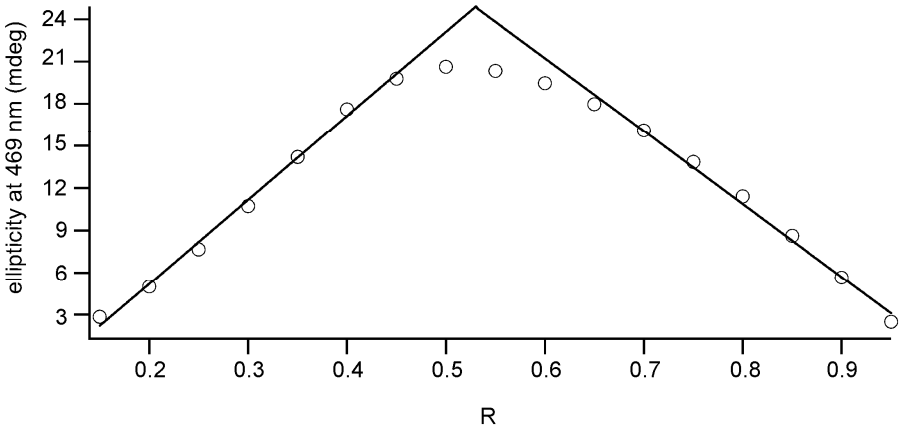


Figure 2.9. Job plot of the continuous fractionation of proflavine into the 2′,5′ dodecanucleotide 5′-(CCGGCCGCGCGC)-2′ and its complement. R = moles of proflavine / (moles of proflavine + moles of nucleic acid binding site). One nucleic acid binding site is defined as four bases or two base pairs. For example the point at R = 0.5 contains 40 μM nucleic acid site and 40 μM proflavine. Buffer conditions and temperature were held constant at 1× BPE, 100 mM NaCl, pH 7 at 25°C [73].

The curvature observed in the Job plot suggests the possibility of more than one binding mode. The curvature can be simulated using the equation,

$$KD_b2 - D_b(1+KN) + (KND_0 - KD_0^2) = 0$$

Where K is the equilibrium association constant in M^{-1}, D_b is the concentration of bound ligand (or concentration of complex formed), and D_0 is the total concentration of ligand. The above equation includes the Job plot requirement that the sum of the total

concentration of ligand and the total concentration of nucleic acid is equal to a constant concentration (*N*).

Equilibrium analysis of the concentration of products reveals that at R = 0.5 the total concentration of proflavine does not always equal the concentration of bound proflavine. Graphically, this effect produces a curved Job plot depending on the equilibrium constant (Figure 2.10A). By this simulation, only a concentration of *ca.* 10^7 or higher would produce a sharp intersection. Additionally, the stoichiometry of ligand-nucleic acid interactions that have equilibrium constants less than *ca.* 10^4 can not be easily determined using this method. Using the experimentally determined *K* for the association of proflavine with 2′,5′ RNA (3.9×10^5 M^{-1}) the curvature observed in the experimentally-determined Job plot is similar to the simulated Job plot with the same *K* (Figure 2.10B). Therefore, the association constant determined by fluorescence is consistent with that which is observed by Job plot analysis.

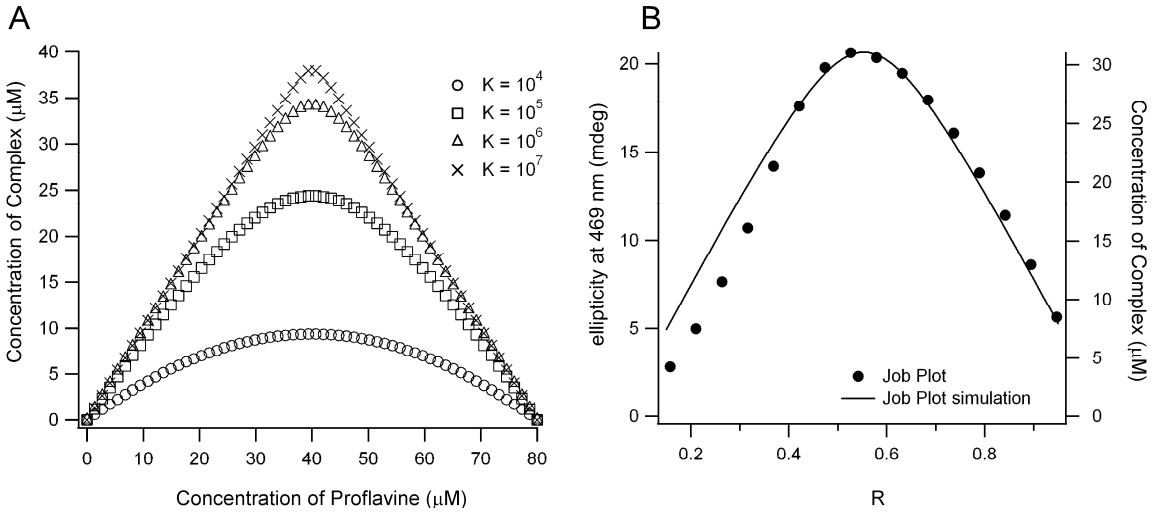


Figure 2.10. Simulation of the curvature observed in Job plots. A) The amount of curvature is dependent on the ligand-nucleic acid association constant. B) Simulation of the amount of curvature expected for an association constant of 3.9×10^5 M^{-1} is consistent with the curvature observed in Figure 2.9.

2.4. CONCLUSIONS

The free energy difference measured for proflavine binding to 2′,5′ RNA versus natural 3′,5′ RNA could result from differences in the duplex structures before or after intercalation or other factors such as differences in hydration. For example, base stacking in a 2′,5′ RNA duplex has been suggested to be less favorable than in 3′,5′ RNA [86], which would make it easier for a 2′,5′ RNA duplex to create an intercalation site. However, an NMR structure of 2′,5′ RNA revealed an A-form helix similar to that of 3′,5′ RNA (albeit with a C2′ *endo* sugar pucker) [87], which suggests that the observed increase in *K* could reflect a difference in the final intercalated states. The observation that ethidium binds 2′,5′ RNA less favorably than standard RNA, while the opposite is true for proflavine, suggests that the pendant ring of ethidium may make less favorable contacts in the minor groove of a locally unwound 2′,5′ RNA duplex than when bound to DNA or RNA. Duplex-specific solvent effects could also contribute to this difference [88,89]. Clearly, explaining the origins of the observed differences in free energies of proflavine and ethidium binding, and verification of the binding mode, require additional investigations. Further thermodynamics are discussed in-depth in Chapter 3.

Our study of ligand-2′,5′ RNA association was the first demonstration of known intercalators binding to a non-3′,5′-linked nucleic acid duplex [73]. Armitage and co-workers have previously reported that hetero-duplexes formed by DNA and RNA with locked nucleic acids (LNA) bind intercalators [90]. However, LNAs possess the same backbone atom connectivity as natural DNA, but with an extra methylene intra-sugar linkage that “locks” sugar conformation.

These results demonstrate the pronounced role of RNA backbone structure in determining the thermodynamics of small molecule binding and also suggest that more significant changes that still maintain the general structure of RNA (*e.g.,* pyranosyl-RNA [91]) could result in even greater binding affinities for intercalators. This ability to enhance small molecule binding by changes in backbone structure could be used to stabilize hybridization of antigene or antisense oligonucleotides [92] and to stabilize intercalator-dependent molecular assemblies [53,58-61,92].

CHAPTER 3

SOLUTION STRUCTURE AND THERMODYNAMICS OF 2′,5′ RNA INTERCALATION[2]

3.1 INTRODUCTION

A large number of planar, polycyclic heterocycles have been shown to intercalate between the base pairs of DNA and RNA helices [91], a mode of binding that can inhibit transcription [93], replication [94], and topoisomerase II activity *in vivo* [95]. Intercalation is one of the two most common modes of non-covalent drug-nucleic acid interaction, with the second mode being groove binding. Nevertheless, even after four decades of continuous study, our understanding of aspects of nucleic acid intercalation remains incomplete [41,72]. For example, the origin of the nearest-neighbor exclusion principle [45], which states that intercalators can bind at no higher stoichiometry than one ligand per two base pairs, remains unresolved [46,51,52,96]. Even specific structural details associated with nucleic acid intercalation remain controversial, including whether an alternating C3′ *endo*/C2′ *endo* sugar pucker along the nucleic acid backbone is a necessary condition for intercalation [32,46-50,74,97].

---

[2] The majority of the work in this chapter was published previously (Horowitz, E.D., Lilavivat, S., Holladay, B.W., Germann, M.W., and Hud, N.V. (2009) "Solution Structure and Thermodynamics of 2′,5′ RNA Intercalation" *J. Am. Chem. Soc.*, DOI: 10.1021/ja810068e)

The energetics of intercalation are influenced by numerous elements of this coupled ligand-binding/structural transition (*e.g.,* electrostatics, dispersive interactions, solvent release, base pair destacking, restriction of conformational entropy) [38]. One necessary and particularly dramatic feature of nucleic acid intercalation is the extending of the double helix at the site of intercalation. It is therefore surprising that, despite numerous physical studies of DNA and RNA intercalation, there is a dearth of data relating to how backbone structure contributes to the thermodynamics and structural transitions associated with intercalation. Intercalation of non-natural backbones could provide fundamental new insights regarding the intercalation of natural nucleic acids. Thus, with the goal of obtaining a deeper understanding of the fundamental principles of nucleic acid intercalation, we have initiated studies of intercalation of non-natural RNA duplexes.

2′,5′ RNA is arguably the closest chemical analog of native 3′,5′ RNA. However, the thermal stability and dynamics of duplexes formed by these two structural isomers are markedly different [98-102]. Thus, we hypothesized duplex 2′,5′ RNA would likely respond differently to intercalation than natural RNA, and therefore represent an excellent first model system for experimentally probing the role of nucleic acid backbone structure in the thermodynamics and ligand specificity of nucleic acid intercalation.

We previously demonstrated that a 2′,5′-linked RNA dodecanucleotide exhibits an association constant, *K*, with proflavine that is 25-fold higher than that of the 3′,5′-linked RNA duplex with the same nucleotide sequence (See Chapter 2) [73]. In contrast, the association constant of the 2′,5′ RNA duplex for ethidium is about half that observed for the 3′,5′ RNA duplex [73]. The structure of the intercalators ethidium and proflavine are

similar. Both are comprised of three fused six-membered rings, both are positively charged at neutral pH, and both have primary exocyclic amine functionalities on the long axis of their fused aromatic ring systems. A clear difference between these two intercalators is the presence of ethyl and phenyl pendant groups on ethidium, which are known to make van der Waals contacts with the minor groove of DNA [74]. The observed difference in the relative affinity of these two molecules for 2′,5′ RNA, compared to both natural DNA and RNA, confirmed our hypothesis that the thermodynamics of intercalation is partially determined by backbone structure, and also emphasized the significant role of the backbone in determining intercalator selectivity.

Although proflavine and ethidium are two well-studied RNA and DNA intercalators [33,41,70], our initial evidence that these molecules bind duplex 2′,5′ RNA (*i.e.,* CD, UV-vis, fluorescence spectroscopy) was not sufficient to conclude that binding occurs through intercalation. Proflavine, for instance, can also bind to the outside of duplex DNA, in addition to intercalative binding [33]. The NMR study presented in this chapter, which was also recently published [103], provides definitive proof that proflavine binds duplex 2′,5′ RNA through intercalation, and suggests that particular backbone structural features are general to nucleic acid intercalation. A more thorough thermodynamic comparison of 2′,5′ and 3′,5′ RNA intercalation also shows the generality of a previously observed enthalpy-entropy relationship for intercalative binding. The potential importance of these results to resolving the elusive origin of the nearest neighbor exclusion principle is discussed.

3.2. EXPERIMENTAL PROCEDURES

3.2.1 Materials

2′,5′ RNA was synthesized using standard phosphoramidite chemistry using the corresponding 3′-TBDMS-protected 2′-phosphoramidites (ChemGenes). Following deprotection, oligonucleotides were separated by length on a 20% denaturing polyacrylamide gel. The band corresponding to the full length product was cut from the gel, extracted from the gel matrix by the crush-and-soak method, loaded onto an anion exchange column, washed with 150 mM NaCl, and eluted with 2.5 M NaCl. The oligonucleotides were then ethanol precipitated, dried, resuspended in dH_2O and passed over a 1 m Sephadex G-10 column. Combined fractions containing the purified product (sodium salt) were pooled, lyophilized, and resuspended in dH_2O. 3′,5′ RNA oligos were purchased from Dharmacon Inc. and used without further purification.

Extinction coefficients for RNA oligonucleotides were determined by strand hydrolysis in 0.3 M NaOH [75] and calculated using reported nucleotide monophosphate extinction coefficients [76]. Extinction coefficients (260 nm) determined for 2′,5′ RNA were: GCCGCGGC, 52,100 M^{-1} cm^{-1}; CCCGCCGCGCCG, 55,000 M^{-1} cm^{-1}; and CGGCGCGGCGGG, 71,600 M^{-1} cm^{-1}. For 3′,5′ RNA: CCCGCCGCGCCG, 52,100 M^{-1} cm^{-1} and CGGCGCGGCGGG, 71,100 M^{-1} cm^{-1}.

Ligand molecules, proflavine (Sigma-Aldrich), acridine orange (Sigma-Aldrich), and ethidium (Fisher Scientific), were purchased and used without further purification.

3.2.2. NMR Spectroscopy

For spectra with exchangeable proton resonances, NMR samples were 1 mM in 2′,5′ RNA duplex, 100 mM NaCl, 30 mM phosphate, pH 6.5, 10% D_2O (500 μL). Spectra were collected at 275 K on a Bruker DRX500 Avance using the 3-9-19 WATERGATE pulse sequence [104]. Imino protons were assigned by 2D NOESY at 282 K. NMR titrations were performed by dissolving dried aliquots of proflavine (20 nmoles each titration) with the NMR sample solution.

For spectra collected in D_2O, the final proflavine-titration sample was lyophilized and resuspended in 250 μL 100% D_2O (Cambridge Isotope Laboratories). 2D NOESY, ^{31}P-decoupled-COSY, ^{31}P-^{1}H HETCOR and TOCSY spectra were collected on a Bruker 600 Avance at 282 K. NOESY spectra were acquired with mixing times of 75, 150 and 250 ms, which were confirmed to be within the linear range of the NOE peak intensity growth. NOE peaks were separated into ‘strong,’ ‘medium,’ and ‘weak’ and were assigned ranges 1.8 to 2.7 Å, 1.8 to 3.3 Å, and 1.8 to 5.0 Å, respectively.

3.2.3. Structure Refinement

Molecular dynamics (MD) simulations were performed using AMBER 8.0 [105], with RNA described by the parm94 force field with the parmbs0 [106] improvements. The 2′,5′ backbone was parameterized by exchanging the standard AMBER RNA charge and atom types for O3′ and H3′ with those of the respective O2′ and H2′. The General AMBER Force Field [107] was used to obtain proflavine parameters. The Jaguar software package (Schrodinger) was used to calculate atomic charges using DFT with a HF/6-31G** basis set at the B3LYP level [108].

For the initial AMBER model, proflavine was manually docked into two intercalation sites that were introduced into an A-form 2′,5′ RNA duplex. Initial inspection of NOE constraints were consistent with proflavine being orientated in the same direction as had previously been observed in the 3′,5′ RNA, *i.e.,* with amino groups pointed towards the major groove [32,47]. As a means to guard against modeling bias, an alternative model was also generated with proflavine rotated 180° with respect to the flanking base pairs (*i.e.,* with exocyclic amines facing the minor groove). Over the course of the simulated annealing, the NOE constraints caused the proflavine to reverse its orientation, thereby supporting our initial orientation of intercalated proflavine.

For annealing simulations, 4-10 ps of high temperature (ranging between 400 – 4000 K) unrestrained MD was used to prepare 300 randomly-generated starting structures, followed by a two-stage structure calculation process. During the first refinement, the molecule was heated to 800 K in the first 5 ps, cooled to 100 K for the next 13 ps, and then cooled to 0 K for the last 2 ps. The temperature of the system was maintained with a varying time constant; 0.4 ps during heating, 4 ps during cooling to 100 K, 1 ps for the final cooling stage, and then reduced from 0.1 to 0.05 for the last picosecond. The force constants for NOE constraints were increased from 3 to 30 kcal mol^{-1} $Å^{-2}$ during the first 5 ps, and then maintained constant for the rest of the simulation. These force constants were applied in the form of a parabolic, flat-well energy term where r is the model distance or torsion angle and k is the respective force constant.

$$E_{constraint} = k(r_2 - r)^2 \quad r_1 \le r < r_2$$

$$E_{constraint} = 0 \quad r_2 \le r \le r_3$$

$$E_{constraint} = k(r_3 - r)^2 \quad r_3 \le r < r_4$$

The values for r_1 and r_4 represent upper and lower distance bounds, defining the linear energetic penalty before and after the flat-well energy term. Distance constraints between base pairs, set at half the constraint strength used for NOE-derived distance constraints, were applied to maintain H-bonding (as indicated by the observation of imino and amino proton resonances). Planarity torsion force constraints of 25 kcal mol^{-1} rad^{-2} were applied to the proflavine molecules throughout the simulations. Planarity torsion force constraints of 200 kcal mol^{-1} rad^{-2}, on N9 of purines and N1 of pyrimidines, were used to maintain the glycosidic bond in plane of the bases. Backbone torsion constraints, including ribose pseudorotation, α, and ζ torsion angles, were held with force constants of 1000 kcal mol^{-1} rad^{-2} based on ^{31}P chemical shift analysis [109]. The backbone angles β ($P_{(i-1)}$-O5′$_{(i)}$-C5′$_{(i)}$-C4′$_{(i)}$), γ (O5′-C5′-C4′-C3′), and ε (C3′-C2′-O2′-P) were not constrained because of insufficient ^1H-^1H and ^{31}P-^1H coupling data, which would define these angles.

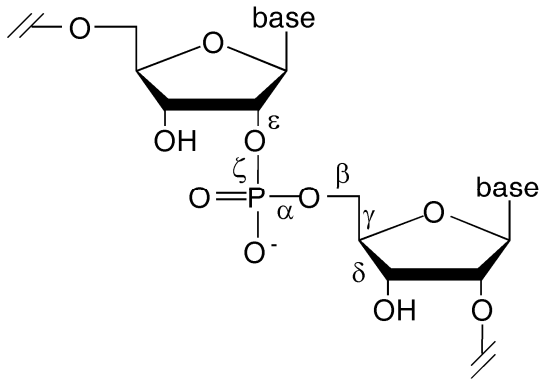


Figure 3.1. Schematic showing the 2′,5′ RNA linkage. The backbone torsion angles are labeled.

The 30 lowest energy structures were each subjected to a second stage of restrained molecular dynamics using the Generalized Born implicit solvent model [110].

The temperature profile, simulation length, constraint forces were identical to those of the first refinement procedure. Ten structures with the lowest energy and NOE violations were selected for the final ensemble and analyzed using VMD [111]. The helical parameters for the 10 lowest energy structures were analyzed using CURVES.

3.2.4. Fluorescence Spectroscopy

Fluorescence measurements were performed on a Shimadzu RF-5301PC spectrofluorophotometer at 298 K. Sample buffer was 1× BPE (8 mM sodium phosphate, 1 mM Na_2EDTA, pH 7) and 100 mM NaCl. Small volumes (*e.g.,* 1 μL) of a concentrated stock of nucleic acid that also contained 1 μM ligand were added to a 200 μL 1 μM solution of ligand. Excitation wavelengths and excitation and emission bandwidths were: proflavine, 455 nm, 1.5 nm, 5 nm; acridine orange, 475 nm, 1.5 nm, 5 nm; ethdium, 510 nm, 5 nm, 10 nm. Equilibrium constants were derived as described by Qu and Chaires [77]. All titrations were performed in triplicate.

3.2.5. Isothermal Titration Calorimetry

ITC measurements were performed using a Microcal VP-ITC at 298 K. For model-free determination of ligand binding enthalpy [112], each injection was 5 μL of 125 μM ligand with the sample cell initially containing 1.4 ml of 250 μM nucleic acid (bp). As opposed to binding site-limited ITC, model-free ITC involves titration of ligand into a high concentration of binding sites. Each titration point produces a similar enthalpy, preventing the need for fitting to a model [112]. Final enthalpy values were determined by fitting a Gaussian curve to histograms of evolved heat per mole of ligand

(corrected for heat of dilution) from multiple injections. Sample buffer was 100 mM NaCl, 1× BPE, pH 7.

3.3. RESULTS

3.3.1. ^{1}H-NMR Confirms Proflavine Intercalation of 2′,5′ RNA

In reference to the imino protons in a self-complementary duplex, there is point of 2-fold symmetry at the center of the duplex (in this case between G4 and C5). This gives rise to four potential imino proton peaks in the ^{1}H-NMR spectrum (Figure 3.2). Solvent exchange as well as end-fraying effects can cause the end imino protons to disappear or appear as broad peaks [113]. Titration of the 2′,5′ RNA duplex $[GCCGCGGC]_2$ with proflavine to a stoichiometry of two proflavine molecules per duplex was monitored by ^{1}H-NMR spectroscopy (Figure 3.3). Over the course of titration, the three original imino resonances disappeared, and four new imino resonances were observed at a proflavine:duplex concentration of 2:1. Additionally, an intermediate set of resonances was observed between these initial and final stages of the titration (Figure 3.3). The coexistence of these intermediate resonances with the initial and final resonances reveals that duplexes with different levels of proflavine loading are in slow exchange on the NMR time scale.

While the ^{1}H NMR spectrum of the ligand-free duplex contains only three imino resonances, more than four imino resonances are associated with the intermediate stages of the titration (most apparent for proflavine:duplex ratios of 0.5:1 and 1:1). This increased number of resonances is consistent with a break in the symmetry of the RNA duplex upon the binding of a single proflavine molecule. The symmetry of the proflavine-

bound duplex is recovered at the last stage, where two proflavine molecules must be bound at two symmetry-related sites.

The four imino protons in the final proflavine titration sample (2:1 proflavine:duplex) were assigned by 2D ^{1}H-NMR. These resonances are shifted upfield from the corresponding imino proton resonances of the unbound (proflavine-free) duplex. An upfield shift in imino proton resonance has been shown to correlate with binding by intercalation, rather than groove binding or outside stacking along the phosphate backbone [114,115].

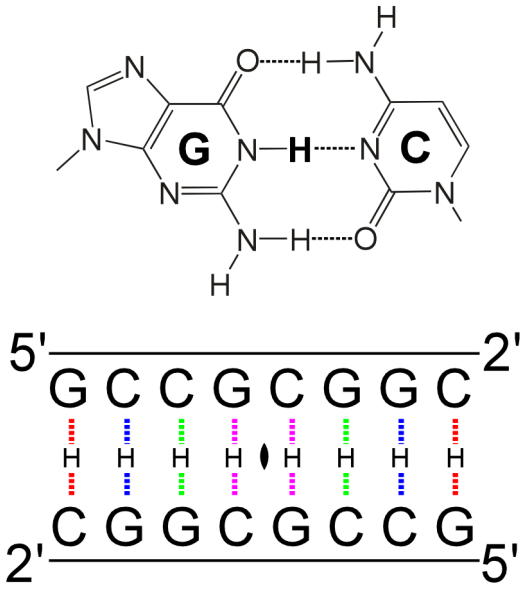


Figure 3.2. Schematic representation of the symmetry of the imino protons in a self-complementary duplex.

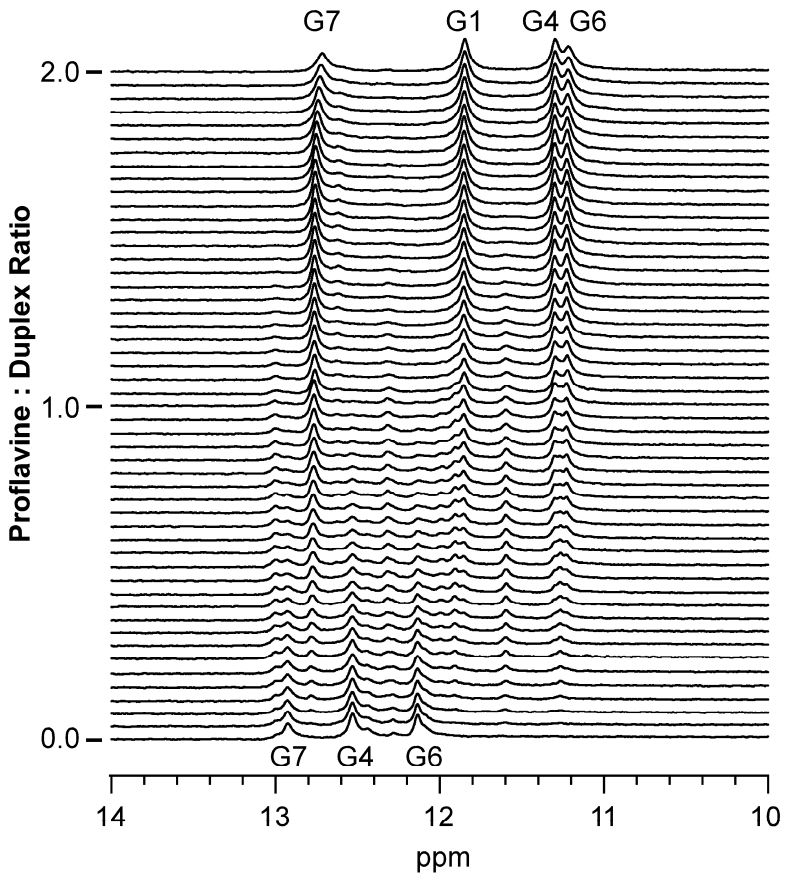


Figure 3.3. ^{1}H-NMR of the imino proton region during the titration of proflavine into a 2′,5′ RNA duplex $[GCCGCGGC]_2$ at 275 K. Duplex concentration was 1 mM, proflavine concentrations ranged from 0 to 2 mM, and are shown to the left of the spectra. Each proflavine addition was in increments of 0.04 equivalents. Proton assignments for the proflavine-free sample (bottom) are from Premraj *et al.* [87]. Proton assignments for the top spectrum are based upon 2D NOESY assignments.

The appearance of a fourth imino proton resonance from the terminal base pairs illustrates that this proton is in slow exchange with the solvent. This reduced rate of exchange could be evidence of reduced fraying of the duplex ends from proflavine stabilization of the duplex [73] or increased end-to-end stacking of proflavine-bound duplexes, which would also protect the terminal imino protons from solvent exchange [116]. Evidence for end-to-end stacking for the proflavine-bound duplex was observed in the form of weak NOEs between G1 and C8.

3.3.2. The Solution Structure of Proflavine-Bound 2′,5′ RNA

The aromatic-H1′ (Figure 3.4) region of a 2D ^{1}H NOESY spectrum, with the intra- and inter-residue aromatic-H1′ NOE "walk-through," [113] is shown in Figure 3.5 for the 2′,5′ RNA duplex with bound proflavine. The G1H8-G1H1′ cross peak exhibits a greater intensity than other comparable NOEs (*e.g.,* compared to G1H8–G1H2′). A similar observation has previously been reported for the 5′-terminal nucleotide of 2′,5′ RNA and 2′,5′ DNA duplexes, and has been attributed to the glycosidic bond being in the *syn* conformation [87,102,117,118]. In the crystal structures of 2′,5′-linked RNA dinucleotides the 5′-*syn* glycosydic bond is stabilized by an intra-residue N3-H-O5′ H-bond [119].

Intercalation causes the local destacking of a base step such that the H1′ proton of the 5′-nucleotide and the aromatic proton of the 3′-nucleotide that flank the intercalation site are separated beyond the limit of the nuclear Overhauser effect (*i.e.,* >6Å) [120,121]. Accordingly, the cross peaks that would correspond to the G4H8-C3H1′ and G6H8-C5H1′ NOEs of a standard duplex are not observed (Figure 3.5). However, the NOE walk-through can be continued through a proflavine resonance assigned to the chemically equivalent H2 and H7 protons. These NOE cross peaks also confirm the binding site of proflavine to the two symmetry-related 5′-CpG-2′ dinucleotide steps (C3pG4 and C5pG6). Interestingly, the same preference for CpG steps, over CpC, GpC, and GpG, steps has long been appreciated for 3′,5′ DNA and RNA duplexes [32,47,74,97,122].

RNA (Table 3.1) and proflavine (Table 3.2) proton assignments were made based on 2D COSY, TOCSY, and NOESY. Twenty-five NOEs were observed between proflavine and the RNA duplex, which define the position of proflavine within the CpG

binding site (Figure 3.6). In the lowest energy structures, proflavine is aligned such that the exocyclic amines face the major groove, similar to the orientation of proflavine in the co-crystal with the 3′,5′ CpG dinucleotide [32,47,97].

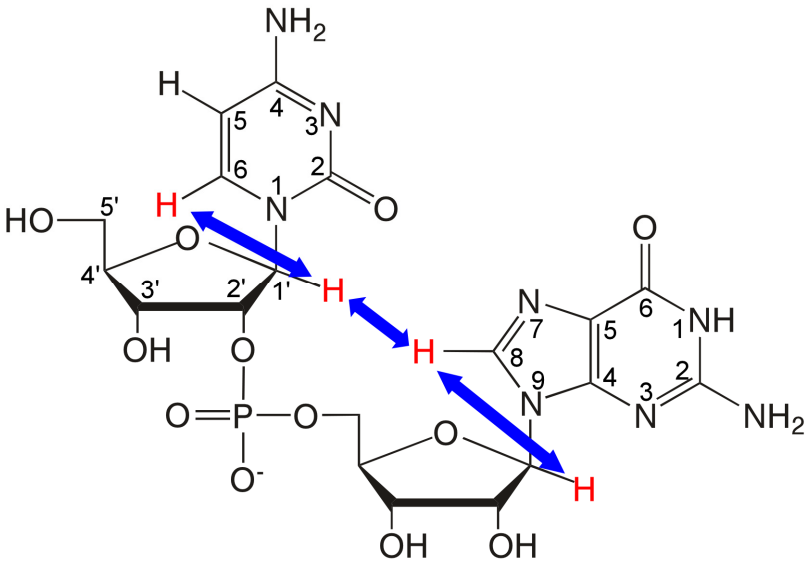


Figure 3.4. Schematic of the NOEs in a dinucleotide step in a 2′,5′ RNA oligonucleotide. Blue arrows represent the NOE peaks used for determining an NOE walkthrough like that shown in Figure 3.5. Atoms are numbered.

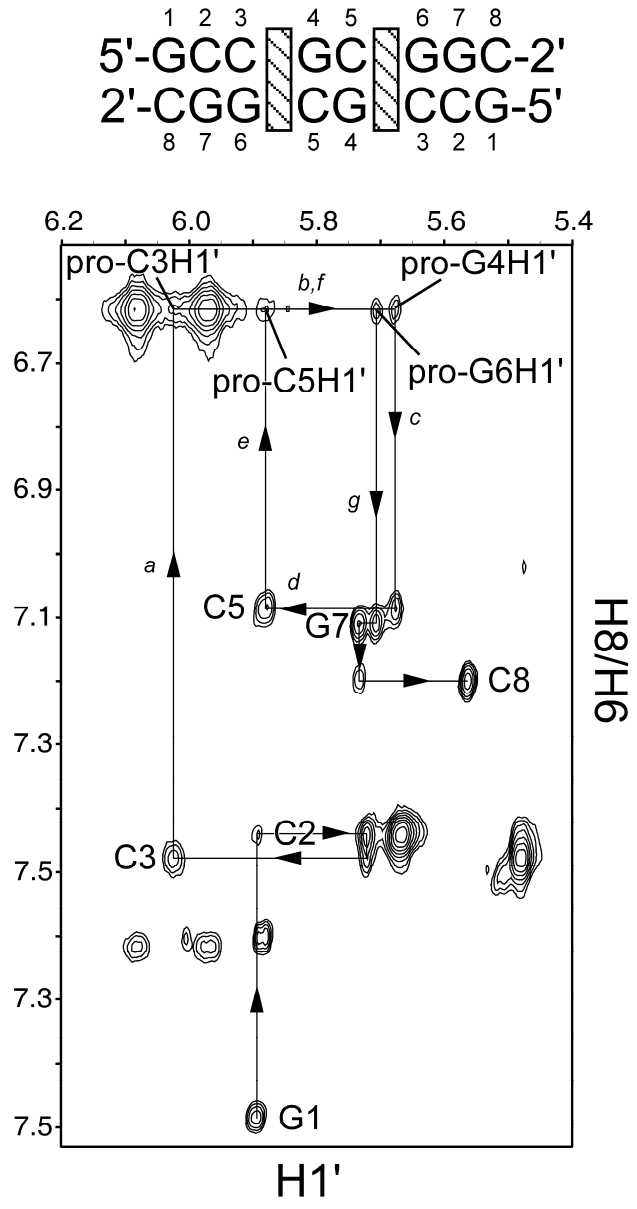


Figure 3.5. The aromatic-H1′ region of a 2D-NOESY of the 2′,5′ RNA duplex $[GCCGCGGC]_2$ with bound proflavine. The aromatic-H1′ NOE walk-through along the duplex is shown. At the intercalation site, the proflavine resonance H2/H7 is used in lieu of the adjacent base. Arrows marked *a* through *g* indicate the proflavine-RNA NOE connectivity. RNA concentration was 2 mM in duplex and proflavine was 4 mM. The spectrum was acquired with a mixing time of 250 ms at 600 MHz in D_2O, 282 K, 60 mM phosphate buffer pH 6.5, 200 mM NaCl.

Table 3.1. Intercalated 2′,5′ RNA ^{1}H Assignments at 282 K (ppm)

	H6/H8	H5	H1′	H2′	H3′	H4′	H5′/H5″	imino
G1	7.89	--	5.89	4.98	4.42	4.15	3.70/2.75	11.87
C2	7.44	5.67	5.72	4.15	4.16	3.49	3.91/3.85	--
C3	7.48	5.48	6.02	4.70	4.47	4.25	4.05	--
G4	7.45	--	5.68	4.64	4.72	4.19	--	11.30
C5	7.09	4.70	5.88	4.48	4.45	4.30	4.11	--
G6	7.39	--	5.71	4.70	4.18	4.16	--	11.22
G7	7.11	--	5.73	4.48	4.36	4.27	4.14	12.72
C8	7.20	5.20	5.56	3.90	3.85	3.99	4.08/3.87	--

Table 3.2. Intercalated Proflavine ^{1}H Assignments at 282 K (ppm)

H1/H8	H2/7	H4/H5	H9
6.08	6.62	5.97	7.62

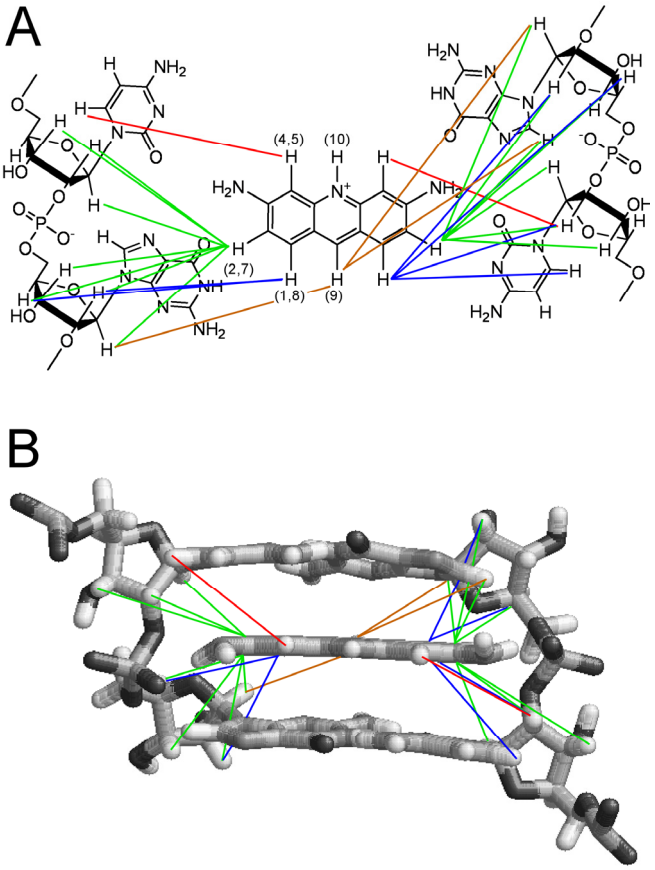


Figure 3.6. A) Diagram illustrating the NOEs observed between proflavine and the CpG steps of the 2′,5′ RNA duplex. The IUPAC acridine numbering is shown around the proflavine ring. B) Diagram illustrating the 3D spatial arrangement of the NOEs shown in A (viewed from the major groove). Structure shown is the lowest energy NMR-derived structure. Colors are used for comparison between A and B.

^{31}P-^{1}H HETCOR heteronuclear spectroscopy was used to assign the seven phosphorous resonances of the 2′,5′ RNA duplex, with and without two molar equivalents of proflavine (Table 3.3). The two ^{31}P resonances associated with the intercalation sites (C3pG4 and C5pG6) are shifted downfield from the unintercalated phosphorous resonances by *ca.* 1 ppm (Figure 3.7). This down-field shift is consistent with what has been previously observed for a BI to BII transition in the DNA backbone [123], and is also associated with intercalation of DNA [124]. The backbone torsion angles ζ and α, corresponding to $C2'_{(i)}$-$O2'_{(i)}$-$P_{(i)}$-$O5'_{(i+1)}$ and $O2'_{(i)}$-$P_{(i)}$-$O5'_{(i+1)}$-$C5'_{(i+1)}$ respectively, influence ^{31}P chemical shift. Correlations have been observed between ζ and α angles and ^{31}P chemical shift [125]. Based upon these previous correlations, the downfield shifts of the phosphorus resonances at the intercalation site correspond to ζ and α angles of 180 ± 60° (*trans*) and -60 ± 60° (-*gauche*), respectively, whereas the phosphorus nuclei at the unintercalated sites have chemical shifts that correspond to ζ and α angles of -60 ± 60° (-*gauche*) and -60 ± 60° (-*gauche*).

Table 3.3. Intercalated and Unintercalated 2′,5′ RNA ^{31}P Assignments at 282 K (ppm)

G1pC2	C2pC3	C3pG4	G4pC5	C5pG6	G6pG7	G7pC8
-0.81	-0.25	0.87	-1.23	0.68	-0.51	-0.57
(-0.86)	(-0.10)	(-0.18)	(-0.61)	(-0.16)	(-0.75)	(0.15)

Values shown in parentheses are the assignments for the unintercalated octanucleotide

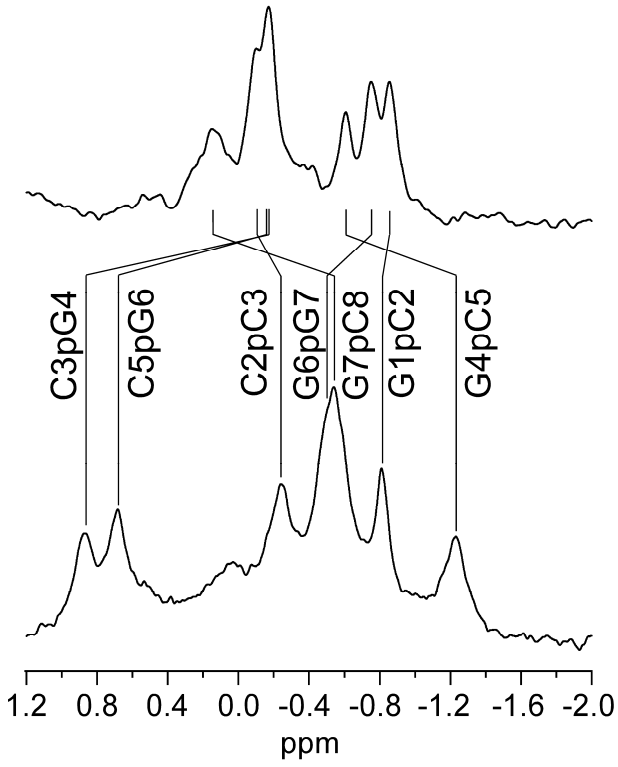


Figure 3.7. ^{1}H-decoupled ^{31}P NMR spectra of the 2′,5′ RNA duplex $(GCCGCGGC)_2$ at 242.9 MHz. Spectra shown are of the unintercalated duplex (Top) and the intercalated duplex containing two proflavine molecules per duplex (Bottom). Spectra were collected at 282 K and processed with 6 Hz line broadening.

Correlations and three-bond coupling constants between H1′ and H2′ resonances, observed using the phase-sensitive ^{31}P-decoupled COSY experiment, were used to determine ribose sugar conformation (Table 3.4). The ribose coupling constants of residues G1, C3, and C5 are consistent with a *South* sugar pucker conformation (*i.e.,* C2′ *endo*), those of G7 and C8 are consistent with a mixed state between *South* and *North* conformations. H1′-H2′ COSY cross peaks were not observed for residues C2, G4, and G6. An absence of H1′-H2′ cross peaks could be caused by small coupling constants, which in conjunction with broad line shapes decreases the intensity of the cross peak. Therefore a lack of a cross peak is indicative of a *North* conformation (*i.e.,* C3′ *endo*) [126]. Thus, at the intercalated 5′-CpG-2′ steps, sugar conformations were *South* for the

5′ residues (*i.e.,* C3 and C5) and *North* for the 2′ residues (*i.e.,* G4 and G6). *This arrangement of sugar puckers around the intercalation site is the converse of that commonly observed in 3′,5′ linked DNA and RNA duplexes*, in which the sugar of the 5′ pyrimidine residue of an intercalation site is typically in a *North* conformation and the sugar of the 3′ purine residue is in the *South* conformation [127].

Table 3.4. Three bond coupling constants for the H1′-H2′ correlations

	^3J H1′-H2′	Sugar Pucker
G1	7 Hz	C2′ *endo*
C2	< 1.5 Hz*	C3′ *endo*
C3	8 Hz	C2′ *endo*
G4	< 1.5 Hz*	C3′ *endo*
C5	6 Hz	C2′ *endo*
G6	< 1.5 Hz*	C3′ *endo*
G7	5 Hz	C2′ *endo*/C3′ *endo*
C8	5 Hz	C2′ *endo*/C3′ *endo*

* Values could not be determined experimentally; the upper limit of 1.5 Hz coupling is based on resonance line width and resolution of 2D spectra.

The number of G1 to C2 and G7 to C8 inter-residue NOEs were about twice the number observed for the other nucleotide steps and, as mentioned above, mixed sugar pucker conformations were observed for G7 and C8. When all constraints for these two terminal nucleotide steps were applied during structure refinement, the terminal base C8 appeared to be held rigidly and in an orientation that was atypical of a terminal duplex nucleotide base (*e.g.,* overly propeller twisted and buckled). We hypothesized that the unusually large number of constraints, and two sugar conformations, were indicative of the terminal base pairs existing in two different conformations, corresponding to the *anti* and *syn* glycosidic bonds of G1. In order to not distort the helix beyond the first base pair

by the simultaneous application of apparently incompatible constraints, the solution state structure was calculated using only the observed G1/C8-C2/G7 inter-nucleotide constraints that were consistent with the *syn* glycosidic bond of G1. Thus, in the final model, the structure of the terminal base pairs only represents one of two conformations that appear to coexist in equilibrium.

The constraints shown in Table 3.5 were used with the AMBER suite of programs to refine a solution state model of the intercalated 2′,5′ RNA duplex. The ten lowest energy structures converged with an all atom RMSD of 0.80 Å (Figure 3.8). When these structures are aligned relative to the proflavine molecules, the base pairs surrounding the intercalation sites are very well defined (Figure 3.8).

Table 3.5. Structural statistics for the 10 lowest energy structures

NOE-derived distance constraints	189
Intra-nucleotide NOEs	119
Inter-nucleotide NOEs	70
Pseudorotation angle constraints	8
Backbone torsion angle constraints	14
H-bond constraints	40
NOE violations > 1.0 Å	0
NOE violations > 0.5 Å	*4.4
RMSD	0.80 Å

The number of constraints shown represents the constraints from only one strand, excluding the H-bond constraints
*The number of violations >0.5 Å were between 3 and 5, the average of the 10 structures is shown

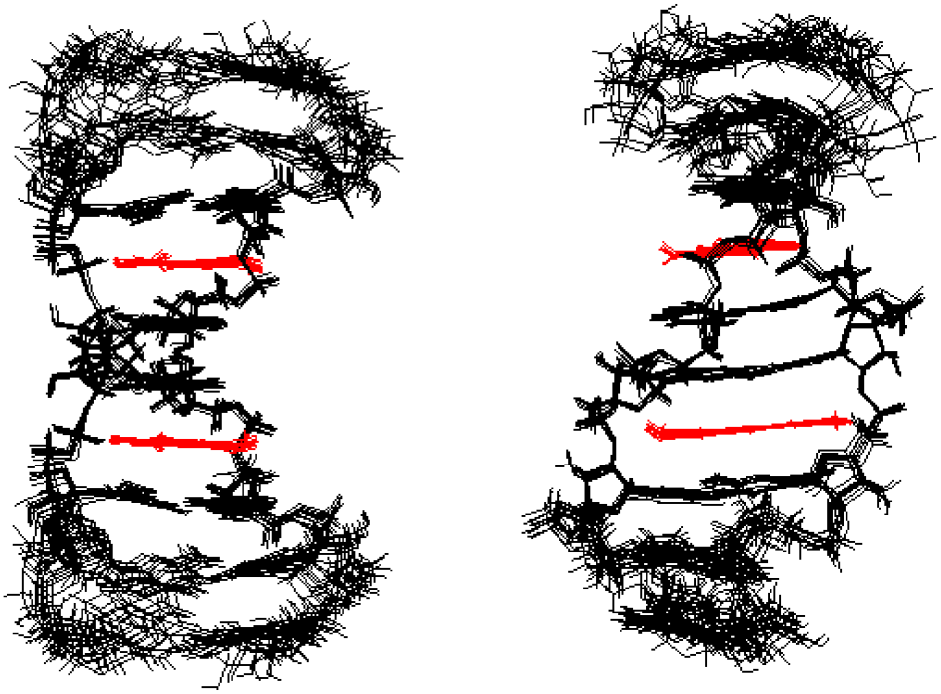


Figure 3.8. Superimposition of the ten lowest-energy NMR-derived structures for the 2′,5′ RNA duplex $[GCCGCGGC]_2$ with bound proflavine. The structures are superimposed with respect to the two proflavine molecules (shown in red). Structures are shown viewing across the minor groove (Left) and into the minor groove (Right). The lowest energy structure was submitted to the Brookhaven Protein Databank (PDB ID *2KD4*).

This structure of a proflavine-intercalated 2′,5′ RNA duplex has a helical twist of only 22º at the CpG steps, while the base-steps flanking the intercalation sites have helical twists of 42, 49, and 42° (C2pC3, G4pC5, and G6pG7 respectively) (Figure 3.9). These helical twist values at unintercalated steps are consistent with those previously observed for the same (unintercalated) 2′,5′ RNA duplex [87]. The 2′,5′ RNA double helix therefore unwinds by an average of 22º at each intercalation site, which is somewhat larger than the unwinding angle observed for proflavine-bound 3′,5′ nucleic acids (*i.e., ca.* 17º) [50]. This difference could be the result of 3′,5′ RNA having a lower

unintercalated helical twist (33º) than 2′,5′ RNA [21], and that the twist of intercalated steps is similar for both forms of RNA.

The helical rise of the base steps neighboring the intercalation sites is larger than that which was observed for the native duplex (*ca.* 3.8 Å vs. 2.5 Å [87]) (Figure 3.9). This observation arises from the unwound intercalation sites causing the neighboring base pairs to be perpendicular to the helical axis. Our NMR model shows that the native helical rise of 2.5 Å increases to 6.6 Å upon intercalation. The increased helical rise at the intercalated base steps is similar to that which is observed for 3′,5′ RNA upon intercalation (*i.e.,* from *ca.* 2.9 Å to 6.8 Å) [21,41,47]. The negative slide values of the unintercalated base steps are consistent with an A-form helix, and are very similar to what was previously reported for unintercalated 2′,5′ RNA [87].

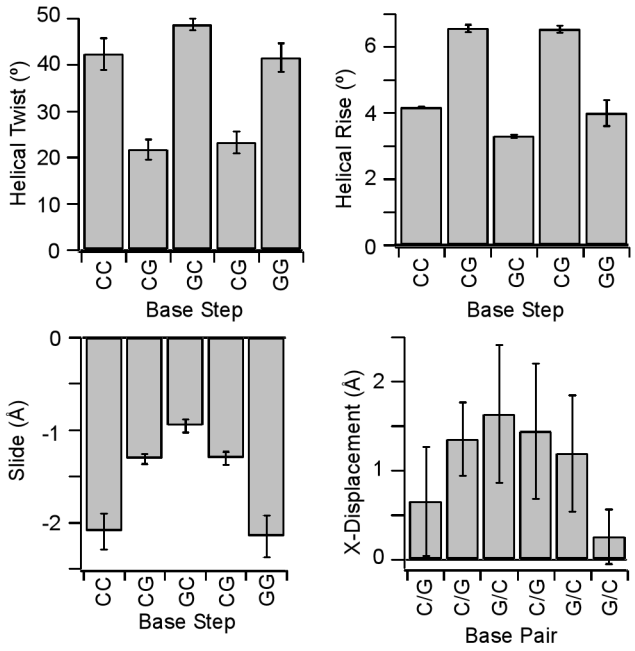


Figure 3.9. Bar graphs of average helical parameters determined by CURVES for the 10 lowest energy structures of the 2′,5′ RNA duplex $(GCCGCGGC)_2$ with proflavine bound at the two CpG steps. Only the central five out of the seven base steps are shown. Error bars indicate standard deviations.

3.3.3. Backbone Conformational Analysis

As mentioned above, a downfield change in ^{31}P chemical shift had previously been correlated with a change in the backbone angle ζ from -60º (*-gauche*) to 180º (*trans*) while α remains constant at -60º (a BI to BII conformational transition), but for native 3′,5′-linked RNA. To test if our observed changes in ^{31}P chemical shift for 2′,5′ RNA were also indicative of a change in the ζ angle, the constraints that had been applied to the backbone angles were removed and the simulated annealing protocol was repeated. Analysis of the resulting ζ angle conformations revealed a very similar grouping of angles as that which resulted from structures refined with the ^{31}P chemical shift-derived ζ angle constraints (*i.e., –gauche*/*trans* for unintercalated and intercalated phosphates respectively) (Figure 3.10). Thus, the distance constraints and sugar pucker constraints derived from the NMR data are consistent with the correlation between ^{31}P Δδ and ζ angle as determined previously for 3′,5′ DNA and RNA. Additionally, when torsional constraints opposite to those derived from ^{31}P chemical shift changes were used (unintercalated phosphates with ζ angles of 180º and intercalated phosphates with ζ angles of -60º) the ζ angles after simulated annealing exhibit a peculiar distribution. First, a narrow distribution of angles is observed near the edge of the window of angles allowed by the applied constraints, implying that the most favorable conformation (as defined by the NMR constraints and the AMBER force-field) is beyond the defined ζ angle allowed rage (± 60º). Second, the resulting RMSD is higher between the lowest energy structures (1.41 Å). These observations strongly support the ^{31}P chemical shift correlations derived for the ζ angles of 3′,5′ DNA and RNA as being applicable to defining the backbone of 2′,5′ RNA [102].

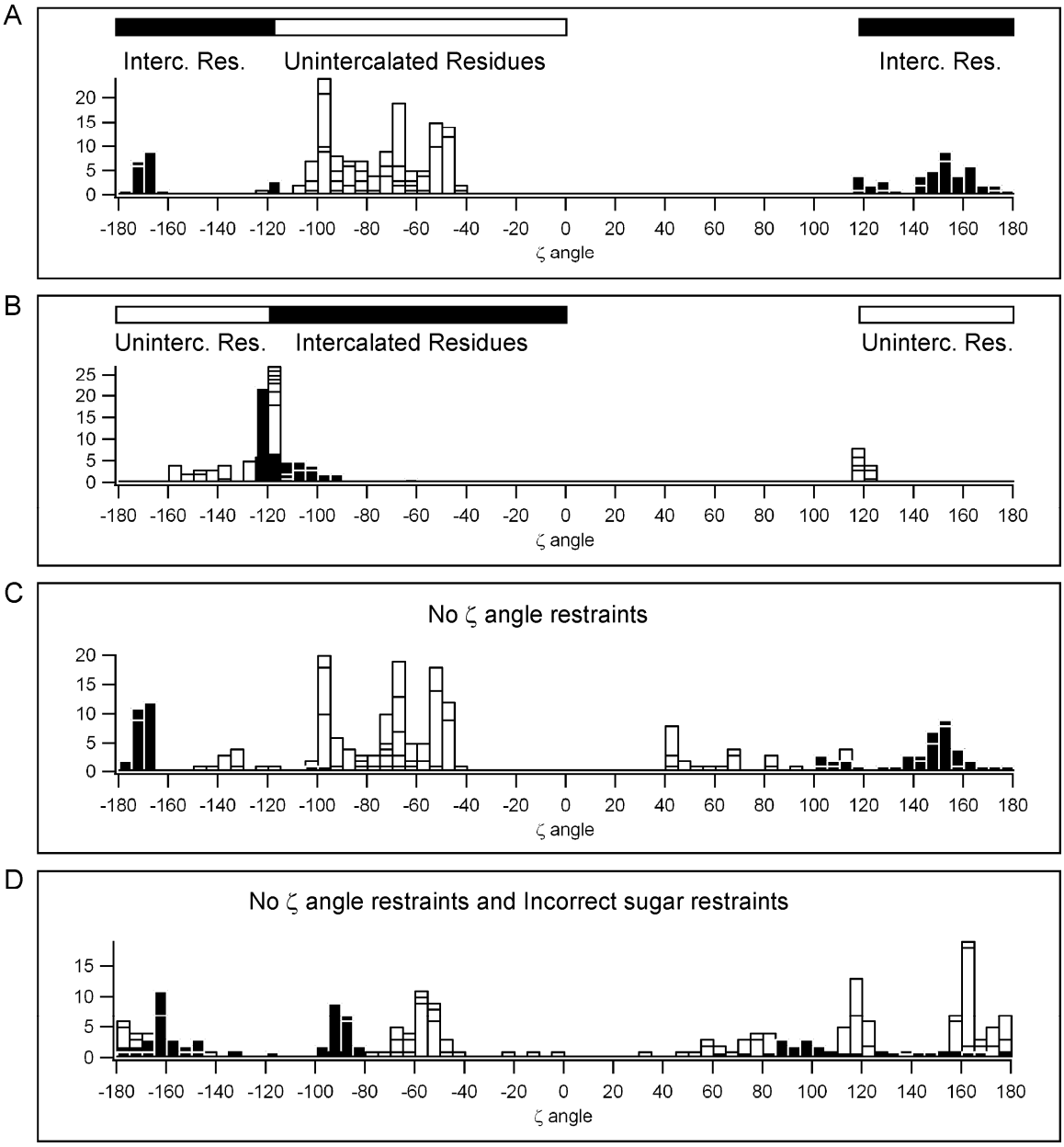


Figure 3.10. Zeta angles compiled from the 30 lowest energy structures after the second annealing refinement. Intercalated residue restraints correspond to ζ angles on the C3pG4 and C5pG6 phosphates. Unintercalated residue restraints correspond to all other ζ angles. A) ζ angles defined by what has been previously observed for natural nucleic acids. B) ζ angles opposite of what was previously observed for natural nucleic acids. C) No ζ angle restraints. D) No ζ angle restraints and the opposite sugar pucker restraints with respect to what is observed on the NMR.

To test if sugar pucker and ζ phosphate angles are necessarily correlated at a site of intercalation in 2′,5′ RNA, a structure minimization was carried out with distance constraints held constant, but with the ζ angle constraints removed and sugar pucker constraints reversed (C3′ *endo* to C2′ *endo*, and visa versa). The resulting ζ angles did not converge to Gaussian distributions as they did when the experimentally-derived sugar pucker constraints were used. These structures also had a higher RMSD (0.99 Å) and more NOE violations greater than 2 Å as compared to no violations greater than 2 Å with the experimentally determined sugar pucker constraints (Table 3.5). This result supports the existence of a structural correlation between sugar pucker and ζ angle at the site of intercalation, implying that C2′ *endo*/C3′ *endo* sugar puckers and a trans (180º) ζ angle are also characteristic of an unwound 2′,5′ RNA helix at the site of intercalation.

3.3.4. Thermodynamics of 2′,5′ RNA Intercalation

Previously, we reported that the association constant of proflavine for duplex 2′,5′ RNA was 25-fold greater compared to duplex 3′,5′ RNA [73]. The fact that 2′,5′ RNA binds proflavine more favorably than 3′,5′ RNA, whereas ethidium was found to bind *less* favorably, indicated that one of these two forms of RNA does not simply bind all intercalators more favorably than the other. To better characterize this difference in binding thermodynamics, we have determined the enthalpy and entropy of proflavine, ethidium, and acridine orange binding to 2′,5′ RNA and 3′,5′ RNA (Figure 3.11).

Using isothermal titration calorimetry (ITC) it was determined that proflavine binds 3′,5′ RNA with a ΔH° of –2.6 kcal mol^{-1}, and 2′,5′ RNA with a ΔH of –13.6 kcal mol^{-1}, for a remarkable difference of -11.0 kcal mol^{-1} (Table 3.6). Given that the ΔG° of

proflavine binding to 2′,5′ RNA is -7.8 kcal mol^{-1} at 25°C, the entropy of binding, calculated as $-T\Delta S° = \Delta G° - \Delta H°$, is +5.8 kcal mol^{-1}. In contrast, the $\Delta G°$ of proflavine binding to 3′,5′ RNA is -5.9 kcal mol^{-1} at 25°C, which implies that $-T\Delta S°$ is -3.3 kcal mol^{-1} (Table 3.6). Thus, although the $\Delta G°$ of proflavine binding to the two forms of RNA differs by only *ca.* 2 kcal mol^{-1}, proflavine binding to 3′,5′ RNA is entropically favored, whereas proflavine binding to 3′,5′ RNA is entropically *not favored* (Figure 3.11).

Acridine orange, which is structurally very similar to proflavine, was also determined to have a favorable binding $\Delta H°$ of -8.7 kcal mol^{-1} and unfavorable $-T\Delta S°$ of +2.3 kcal mol^{-1} for binding to 2′,5′ RNA. The difference in acridine orange binding to 3′,5′ RNA follows that same trend as proflavine (*i.e.,* more favorable enthalpy and less favorable entropy for 2′,5′ RNA) but the difference in these thermodynamic parameters is not as great as that observed for proflavine (Figure 3.11). In contrast to proflavine and acridine orange, the enthalpy and entropy of ethidium binding does not vary appreciably between the two RNA isomers. For these three intercalators, a plot of entropy versus enthalpy of binding, as described by Chaires for small molecule DNA ligands [128], reveals a trend in enthalpy-entropy compensation that appears to be universal for simple intercalators. While the source of this apparent enthalpy-entropy compensation for DNA binding is not revealed by such analysis, it nevertheless appears that this trend is likewise followed by the thermodynamics of 2′,5′ RNA intercalation (Figure 3.12).

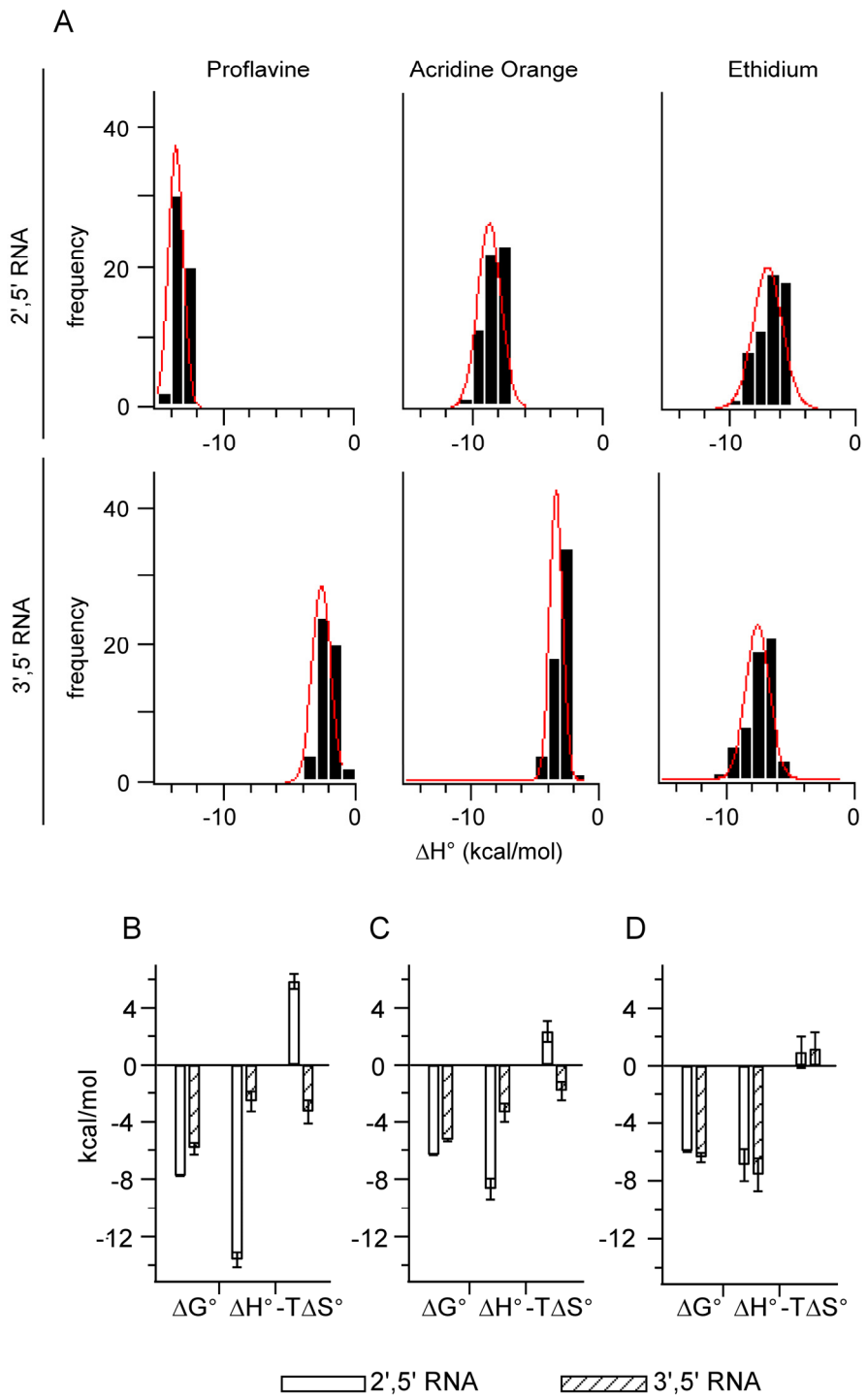


Figure 3.11. A) Distributions of enthalpy values determined by model-free ITC. Gaussian fits to the histograms are shown in red. Thermodynamic values for the binding of three known intercalators to 2′,5′ RNA and 3′,5′ RNA at 25°C. B) proflavine, C) acridine orange, and D) ethidium. Error bars represent standard deviations. Enthalpy values were determined by ITC. Free energy values were determined by both ITC and fluorescence-monitored titration.

Table 3.6. Free energy (ΔG°), enthalpy (ΔH°) and entropy (-TΔS°) values for the intercalation of 2′,5′ RNA and 3′,5′ RNA

	$\Delta G°_{2',5'}$	$\Delta G°_{3',5'}$	$\Delta H°_{2',5'}$	$\Delta H°_{3',5'}$	$-T\Delta S°_{2',5'}$	$-T\Delta S°_{3',5'}$
Proflavine	-7.8	-5.9	-13.6	-2.6	+5.8	-3.3
Acridine Orange	-6.4	-5.3	-8.7	-3.4	+2.3	-1.9
Ethidium	-6.0	-6.4	-6.9	-7.6	+0.9	+1.2

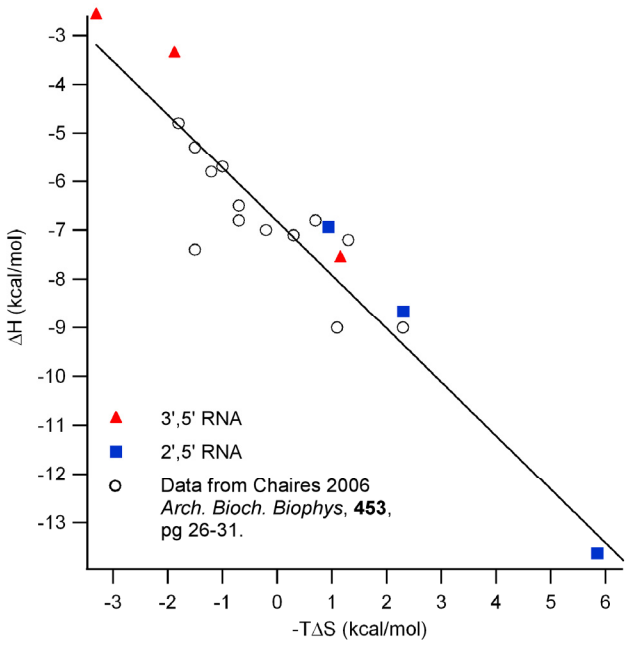


Figure 3.12. The linear relationship between enthalpy and entropy for intercalators binding to nucleic acids. The compilation of data reproduced from the work of Chaires [128] shows the linear relationship for intercalators binding to DNA. The overlyed line is a non-linear least squares fit of all of the data shown. The same trend is observed for 3′,5′ RNA as well as 2′,5′ RNA.

3.3.5. CD Spectra of Intercalated 2′,5′ RNA

We previously noted that the induced CD bands of proflavine (400-500 nm) observed upon binding to 2′,5′ RNA are of almost the same intensity as the nucleic acid bands observed between 225 and 300 nm [73], which is more intense than the induced bands typically observed upon intercalation in 3′,5′ RNA or in DNA. The induced CD bands of proflavine when bound to 2′,5′ RNA are also biphasic-a trend observed for

groove binding small molecules [84]. These observations gave reason to question whether the initial evidence of proflavine binding to 2′,5′ RNA was indeed indicative of intercalation.

In Figure 3.13, CD spectra of the proflavine-bound 2′,5′ RNA octanucleotide used for the NMR experiments (spectrum taken of actual NMR sample), compared to spectra of the previously published 2′,5′ RNA dodecanucleotide with proflavine, as well as a 3′,5′ RNA dodecanucleotide with bound proflavine [73]. The 2′,5′ RNA octanucleotide used in the NMR studies reported above also gave rise to a biphasic induced proflavine band of similar intensity to the nucleic acid bands. Large biphasic ICD signals have been previously demonstrated to arise from ligands stacking along the outside of the helix [84]. The unpredictable size of the intercalated proflavine ICD band demonstrates that, while analysis of the size and shape of the CD spectra of a small molecule bound to DNA can suggest a particular binding mode, such interpretations are potentially misleading. As is shown here, small molecules bound to nucleic acids with different backbone linkages even as similar to native RNA as 2′,5′-linked RNA can produce spectra that are deceiving in terms of ‘rule of thumb’ analysis.

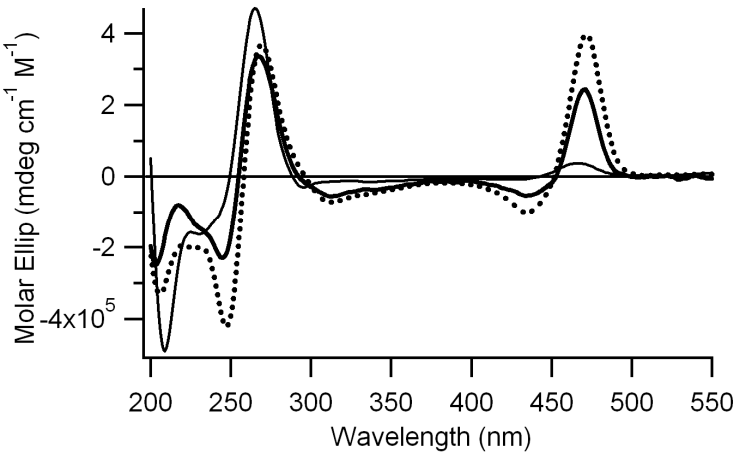


Figure 3.13. CD spectra of the 2′,5′ RNA dodecanucleotide duplex, (CCCGCCGCGCCG)·(CGGCGCGGCGGG) (dotted line, 6.67 μM duplex, 20 μM proflavine), the comparable 3′,5′ RNA dodecanucleotide (solid line, 6.67 μM duplex, 20 μM proflavine), and the 2′,5′ RNA NMR sample duplex $[GCCGCGGC]_2$ (bold solid line, 2 mM duplex, 4 mM proflavine) at 282 K.

3.4. DISCUSSION

3.4.1. Structural Features of 2′,5′ RNA Intercalation

For 3′,5′ RNA and DNA duplexes intercalation is most favored at Py-p-Pu steps [46]. This preference has been attributed to the relatively poor base stacking at Py-p-Pu steps [116,129,130], which makes base-*destacking* upon intercalation *less unfavorable* than at other steps. The structure reported here of proflavine-intercalated 2′,5′ RNA provides direct evidence that, for the sequence studied, Py-p-Pu steps are also preferred sites of intercalation. It has been suggested that all base stacking within a 2′,5′ RNA duplex is less favorable than that in a 3′,5′ RNA duplex [99,131]. It was therefore not obvious at the start of this investigation whether the CpG step would be the most favored site for intercalation by proflavine.

3.4.2. Inter- and Intramolecular H-bonding in the Proflavine-2′,5′ RNA Complex

The proflavine-2′,5′ RNA structure reported here revealed H-bonds between the amino groups of proflavine and phosphate oxygens (H-O distances of *ca.* 1.9 Å). These interactions likely contribute to the comparatively high association constant between proflavine and 2′,5′ RNA. In addition, H-bonds between the 3′-OH and phosphate oxygen atoms exist at every base step. This structural feature was also noted by Premraj *et al*. [87], and we have found that this intra-backbone H-bonding of a 2′,5′ RNA duplex is maintained even at intercalated steps.

3.4.3. 2′,5′ RNA Intercalation Structural Requirements

Voet proposed, based on the observations of Jain and Sobell [132,133] on ethidium-bound dinucleotide crystal structures, that the necessity of alternating sugar puckers in an intercalated duplex could provide an energetic barrier that restricts intercalation to one ligand binding between every-other base pair (*i.e.,* the nearest-neighbor exclusion principle) [46]. However, exceptions to this rule have been observed and attributed to variations in H-bonding between the intercalating molecule and the phosphate backbone [47,97]. Based upon computational studies, Kollman and coworkers noted that the alternating of C3′ *endo*/C2′ *endo* (*North/South*) sugar pucker conformations around an intercalation site is energetically more favorable, but ligand interactions with the phosphate backbone can also cause different sugar puckers to exist [49]. Although more recent calculations regarding the nearest-neighbor exclusion principle and intercalation energetics have been conducted with more current force-field parameters

[52,134,135], the energetics of alternating sugar puckers at intercalation sites has apparently not since been reinvestigated.

The sugar puckers around a 2′,5′ RNA intercalation site were found here to alternate between *South* and *North* conformations for the pyrimidine and purine bases, respectively. This observation is exactly the opposite of the trends discussed above observed in natural RNA. At first glance this difference seems to be very large. However, Premraj *et al*. noted that, while (unintercalated) 2′,5′ RNA and 3′,5′ RNA helices have different sugar puckers (C2′ *endo* and C3′ *endo,* respectively), the two forms of RNA maintain a similar phosphate-phosphate distance of 5.9 Å. These authors termed this state of the RNA backbone the *compact form* [118,131]. The sugar pucker associated with the compact form is in contrast to the sugar pucker of the *extended form* pucker (C3′ *endo* for 2′,5′ RNA and C2′ *endo* for 3′,5′ RNA), which places the phosphate-phosphate distance at 7-7.5 Å [102,136]. Using this terminology, the sugar pucker geometries at the intercalation site in 2′,5′ RNA are compact/extended (Py/Pu) and therefore are actually the same, or at least analogous, to that which has been observed for 3′,5′ RNA, which is also compact/extended (Pyr/Pur) at an intercalation site.

When the opposite arrangement of sugar puckers are forced on the intercalation site, during structure refinement, the angle ζ is found to rotate into unfavorable conformations that are less well defined, thereby increasing the RMSD of the lowest energy structures and the number of NOE constraint violations. This observation indicates that the extended conformation of the purine nucleoside and the unwinding of the phosphate angle ζ allow for the formation of an intercalation site, therefore requiring these structural features for proflavine-2′,5′ RNA intercalation. In addition, intercalation

at the CpG steps is consistent with the recent observation that CpG steps prefer to be in the unwound BII conformation (*trans* ζ) in DNA [137]. The common observation of the compact/extended sugar pucker conformations, as well as the *trans* ζ angle conformation in intercalated 3′,5′ DNA and RNA, lends support to the proposal that these features are general structural requirements for intercalation of a nucleic acid with a phosphate-ribose backbone. Exceptions to this case may include intercalators that bind in perpendicular orientations to the base pairs, such as daunomycin, in which case there is significant base pair buckling and less duplex unwinding [96]. Additionally, the extensive non-stacking interactions of molecules like daunomycin to 3′,5′ nucleic acids may induce a homogeneous compact/compact sugar pucker conformation [47,97].

3.4.4. 2′,5′ RNA Intercalation Thermodynamics

It was surprising to observe such a large difference between the thermodynamics of proflavine binding to 2′,5′ RNA versus 3′,5′ RNA, especially given that no appreciable difference was observed for ethidium binding. H-bonding by the exocyclic amines of proflavine to phosphate oxygens could be a source of significant binding enthalpy in the 2′,5′ RNA complex. The positive charge of proflavine, often drawn only on the protonated N10 nitrogen, is likely delocalized such that partial positive charge resides on the exocyclic amines. The resulting charge-charge interactions between these amino groups and the anionic phosphate oxygens in the 2′,5′ RNA complex would provide a larger enthalpic contribution to binding than typical H-bonds (*e.g.,* water-phosphate H-bonds) [138]. Although proflavine has also been shown to form amine-phosphate oxygen H-bonds in complex with 3′,5′ RNA [47,97], the full enthalpy of these interactions may

not be evident in the *net* enthalpy of proflavine binding. Apparently in order for the 3′,5′ RNA backbone to participate in these H-bonds, it must maintain all *North* sugar puckers around the intercalation site, and not the more energetically favored alternating *North-South* pattern of sugar puckers [49]. In contrast, the intercalation-favored *South-North* pattern of sugar puckers is compatible with the formation of these H-bonds in the 2′,5′ RNA-proflavine complex. The more favorable enthalpy of proflavine binding to 2′,5′ RNA may also reflect a greater difference in the enthalpy of stacking interactions of the intercalated and unintercalated states of 2′,5′ RNA, as compared to the two analogous states of 3′,5′ RNA.

Ethidium presumably has an equivalent capability to form H-bonds with the 2′,5′ RNA backbone and to stack with the bases, but thermodynamic measurements reveal similar binding enthalpy of ethidium to both forms of RNA. It is possible that ethidium also participates in more favorable H-bonding and base stacking with 2′,5′ RNA than 3′,5′ RNA, but that these enthalpic gains are balanced by equally destabilizing interactions. For example, it was noted by Premraj *et al.* that the minor groove width of the 2′,5′ RNA helix is about 1 Å less than that of 3′,5′ RNA [131]. This structural difference could severely disrupt the minor groove packing interactions observed for the phenyl and ethyl of intercalated ethidium in the 3′,5′ RNA crystal structure [74].

Acridine orange represented an opportunity to probe how the thermodynamics of binding would change if the H-bonds between proflavine and 2′,5′ RNA were disrupted. Consistent with this expectation, the favorable ΔH° of proflavine binding to 2′,5′ RNA decreases from -11.0 kcal mol^{-1} to -5.3 kcal mol^{-1} when the exocyclic amines are methylated. It should be noted however, that the binding free energy of acridine orange to

2′,5′ RNA might also be less favorable compared to proflavine because acridine orange has a self-association constant in water that is an order of magnitude higher than that of proflavine [139,140], although little difference is observed between the thermodynamics of proflavine and acridine orange binding to 3′,5′ RNA.

3.5. CONCLUSIONS

We have shown that proflavine binds in the same orientation with the same nucleotide step preference to 2′,5′ and 3′,5′ nucleic acids. However, when proflavine intercalates in 2′,5′ RNA there is a larger enthalpic gain ($\Delta\Delta H^\circ$ = -11 kcal mol^{-1}) compared to intercalating in 3′,5′ RNA, which could be due to more favorable H-bonding (charge-charge) and increased net stacking interactions in 2′,5′ RNA.

We have also identified several structural features that are common between intercalated 2′,5′ RNA and 3′,5′ RNA duplexes. These common features include unwinding angle, change in helical rise, and the compact/extended sugar pucker motif at the intercalation site. Our results suggest that two structural transitions are coupled in the creation of an intercalation site: ζ rotates from the *–gauche* conformation to the *trans* conformation and the purine at the 2′ end of the binding site changes from a *South* (compact) to a *North* (extended) conformation. The data presented here is thus consistent with the proposal of Kollman and coworkers that the observed alternating sugar pucker is the most energetically favorable conformation [49], even with RNA containing a different backbone connectivity.

As the single example presented here illustrates, ligand binding to nucleic acids with an alternative backbone provides fresh insights regarding the structural requirements

for intercalation and the thermodynamics of ligand binding. Additional studies of ligand interactions with modified nucleic acids could further facilitate our understanding of the molecular recognition of natural nucleic acids. A more in-depth understanding of such interactions could also facilitate the development of molecular systems that allow protein-free template-directed synthesis [54,61] and intercalation-driven nanostructures [141] that utilize Watson–Crick base pairs as well as non-canonical base pairs [58-60,142]. Finally, even determining the long-elusive origin of the nearest-neighbor exclusion principle appears obtainable through the use of modified nucleic acids.

CHAPTER 4

INTERCALATION AS A SOLUTION TO THE STRAND CYCLIZATION PROBLEM IN A PRE-RNA WORLD

4.1. INTRODUCTION

Over the past twenty years, significant evidence has been presented in support of the 'RNA world' hypothesis, which proposes that RNA polymers predated coded proteins in early life [143,144]. Current support for this hypothesis includes the fact that contemporary life still uses RNA as both an informational polymer and in chemical catalysis [11]. The ability of RNA to catalyze reactions is exemplified by the identification of natural and artificial ribozymes that promote a wide variety of chemical reactions [145], as well as the fact that the catalytic core of the ribosome is comprised of RNA [11]. Despite the attractiveness of the RNA world as a hypothetical stage of early life, it is not clear how RNA (or a proto-RNA predecessor [66,68,146-149]) would have been initially synthesized without the aid of protein enzyme catalysis.

Several distinct proposals have been presented for the abiotic origin of the first RNA-like polymers [53,64,149-153]. Perhaps the most notable is that of Ferris and coworkers, in which mineral surfaces are used to locally concentrate and promote the polymerization of ligation-activated mononucleotides, an approach that allows formation of single-stranded RNA strands up to around 50 nucleotides in length [154]. However, it is not clear how mineral surfaces would have selectively polymerized only those mononucleotides that can form Watson–Crick base pairs, from what would have likely been a complex mixture of closely-related molecules in the prebiotic chemical inventory

[54]. Earlier work by Orgel and coworkers had demonstrated that oligonucleotides (of rather restricted nucleotide sequence) can serve as templates for the polymerization of activated mononucleotides in solution, resulting in duplexes with Watson–Crick base pairs [155,156]. However, it still remains an open question regarding how the first gene-length duplex polymers of RNA (or proto-RNA) would have assembled from mononucleotides or even short oligonucleotides without pre-existing templates.

Oligonucleotide cyclization is a formidable problem that most proposals for the abiotic synthesis of RNA polymers must face. Specifically, short ligation-activated oligonucleotides (*i.e.,* di- to octanucleotides) undergo efficient intramolecular ligation, which greatly limits polymer growth [54,157-159]. Chemists have long appreciated that the length of polymers formed by stepwise, kinetically-controlled coupling of monomers or oligonucleotides can be greatly limited by intramolecular chain termination (*i.e.,* cyclization) [160]. When a nascent polymer becomes sufficiently long to sample conformations allowing intramolecular bond formation, polymer growth ceases. The length at which a growing polymer cyclizes depends largely on the rigidity of that polymer – often described as its persistence length. The persistence length of single stranded nucleic acids is around three nucleotides [161], whereas the persistence length of a duplex with Watson–Crick base pairs is on the order of 150 bp [162]. Nucleic acids must be at least *ca.* five residues in length to efficiently form duplexes, even at temperatures near the freezing point of water and moderately high oligonucleotide concentrations (10^{-4}-10^{-3} M). Thus, given a pool of chemically activated di-, tri- and tetranucleotides, one would predict that such short oligonucleotides would cyclize efficiently. This prediction is borne out by experiment [157-159] including model

prebiotic reactions where activated mononucleotides are condensed on a mineral surface [64]. Furthermore, activated hexanucleotides in solution have been shown to primarily form cyclic products even if their sequences have the potential to form duplexes with only G·C base pairs [158]. Clearly, strand cyclization would have inhibited the prebiotic production of RNA-like polymers, unless a mechanism existed to increase the persistence length of oligonucleotides during polymerization.

We report that molecules which intercalate the base pairs of nucleic acid duplexes can circumvent the strand cyclization problem. Specifically, we demonstrate that certain intercalators (or 'molecular midwives' [53]) promote the coupling of activated tetranucleotides into long duplex polymers, while only short cyclic oligonucleotides are formed in the absence of an intercalator. We also show that intercalator-mediated polymer formation is ligand- and base pair-specific; size matching is required between the intercalator and the base pairs of a duplex. These data thereby lend support to the hypothesis that nucleic acid structure, including base pairing structure, may have originally been selected for and templated by small molecules that intercalate base pairs [53,54].

4.2. EXPERIMENTAL PROCEDURES

4.2.1. Materials

Oligonucleotides were purchased from Oligos Etc. (Wilsonville, OR) and HPLC purified before use. Ethidium bromide (Fisher Scientific), proflavine hemisulfate (Sigma), and coralyne chloride (Sigma) were used as received. aza3 was synthesized as

previously reported [163]. N-cyanoimidzaole was purchased from Toronto Research Chemicals (North York, ON).

4.2.2. Watson–Crick Chemical Ligations

Ligation reactions were 200 μM in oligonucleotide strand, 10 mM triethylammonium MES (pH 6), 5 mM $MnCl_2$. For reactions containing ethidium, the intercalator was added as indicated in figure legends. After 15 min equilibration at 4 °C, N-cyanoimidazole (from an H_2O stock) was added to 250 mM. Ligation reactions were incubated up to 72 h at 4ºC. The reaction products were then ethanol precipitated and resuspended in PAGE loading buffer. Reactions products were separated using denaturing (7 M urea) PAGE (19% acrylamide monomer: 1% bis acrylamide) with TBE running buffer. These gels were then stained with SYBR Gold (Invitrogen).

4.2.3. Homo-Adenine Chemical Ligations

Ligation reactions were 500 μM in $d(pA_6)$, 10 mM triethylammonium MES (pH 6), 5 mM $MnCl_2$. Reactions with intercalators were 750 μM in intercalator. Ligation reactions were incubated up to 24 h at 4ºC. Reactions products were separated using denaturing (7 M urea) PAGE (19% acrylamide monomer: 1% bis acrylamide) with TBE running buffer. These gels were then stained with SYBR Gold (Invitrogen).

4.2.4. HPLC Separation of Reaction Products

Reactions were 200 μM in tetranucleotide, 5 mM $MnCl_2$, 10 mM MES-TEA (pH 6) and 600 μM ethidium (for the reaction containing the intercalator). N-cyanoimidazole

was added to 25 mM from a stock solution, and the reactions were held at 4 °C. At each time interval, a 10 μL aliquot was removed and diluted in 90 μL 22 mM EDTA to quench the reaction. Aliquots were then immediately injected onto an Agilent 1100 with a 4.6 mm x 250 mm Phenomenex Luna C18 at ambient temperature. The gradient conditions were: Solvent A=100 mM triethylammonium acetate, pH 7. Solvent B=Acetonitrile. 0-12 min, 7.5% B. 12-20 min, 7.5-20% B. 20-25 min, 20-70% B. 25-30 min, 70% B.

4.2.5. ^{1}H NMR of Tetranucleotide Assembly

The tetranucleotide d(pCGTA) at 200 μM in strand is dissolved in buffer containing 10 mM MES-TEA (pH 6) and 5 mM $MgCl_2$ with 10% D_2O. Ethidium titration was performed by adding dry ethidium to the solution to adjust the concentration from 0 to 600 μM ethidium. Spectra were collected at 280 K on a Bruker DRX500 Avance using the 3-9-19 WATERGATE pulse sequence [104].

4.2.6. UV-vis Thermal Denaturation of Assemblies

UV-vis spectra were acquired on a Hewlett Packard 8453 diode array spectrophotometer equipped with an Agilent 89090A Peltier temperature controller. UV melting profiles were produced by monitoring absorbance at 260 nm while increasing the temperature at a rate of 1ºC min^{-1} from 5 to 95ºC. Samples were prepared with and without ethidium (600 μM) and contained 200 μM tetranucleotide, 10 mM MES-TEA (pH 6), and 5 mM $MnCl_2$.

4.3. RESULTS

4.3.1. Intercalators Prevent Strand Cyclization

A number of molecules are known to intercalate nucleic acid duplexes, such as the fluorescent dye molecules commonly used to detect and visualize DNA [164]. We have previously demonstrated that the free energy associated with duplex intercalation can promote the protein enzyme-free ligation of oligonucleotides [54,61]. To determine if intercalation-mediated assembly can also circumvent the strand cyclization problem (Figure 4.1A), we investigated the chemical ligation of the tetranucleotide d(pCGTA) as a model system. The sequence of this oligonucleotide is such that multiple copies, if assembled into a duplex, are capable of forming long, regularly nicked Watson–Crick duplexes, with the 5′ and 3′ ends of the nick sites in close proximity for chemical ligation (Figure 4.1B).

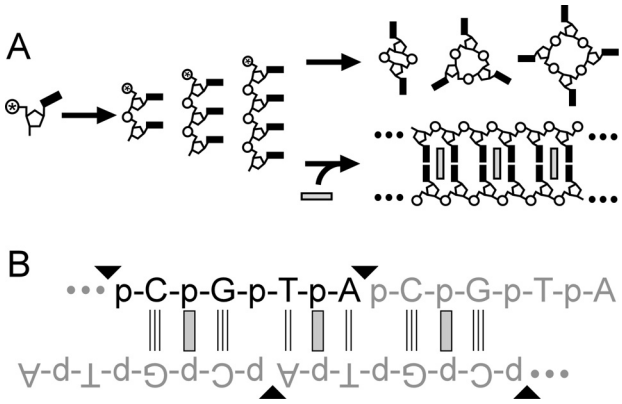


Figure 4.1. A) Schematic illustration of the strand cyclization problem (top route) and the proposed solution by intercalation (bottom route). In the absence of intercalators, the expected products of chemically-activated di-, tri- and tetra-nucleotides are cyclic oligonucleotides. In the presence of intercalators (shown as grey rectangle), the predicted products are long duplex polymers. B) The extended, regularly nicked duplex resulting from the assembly of multiple copies of the tetranucleotide d(pCGTA) upon intercalation. Triangles indicate sites of backbone nicks, which can be closed by condensation after chemical activation of the terminal phosphate group.

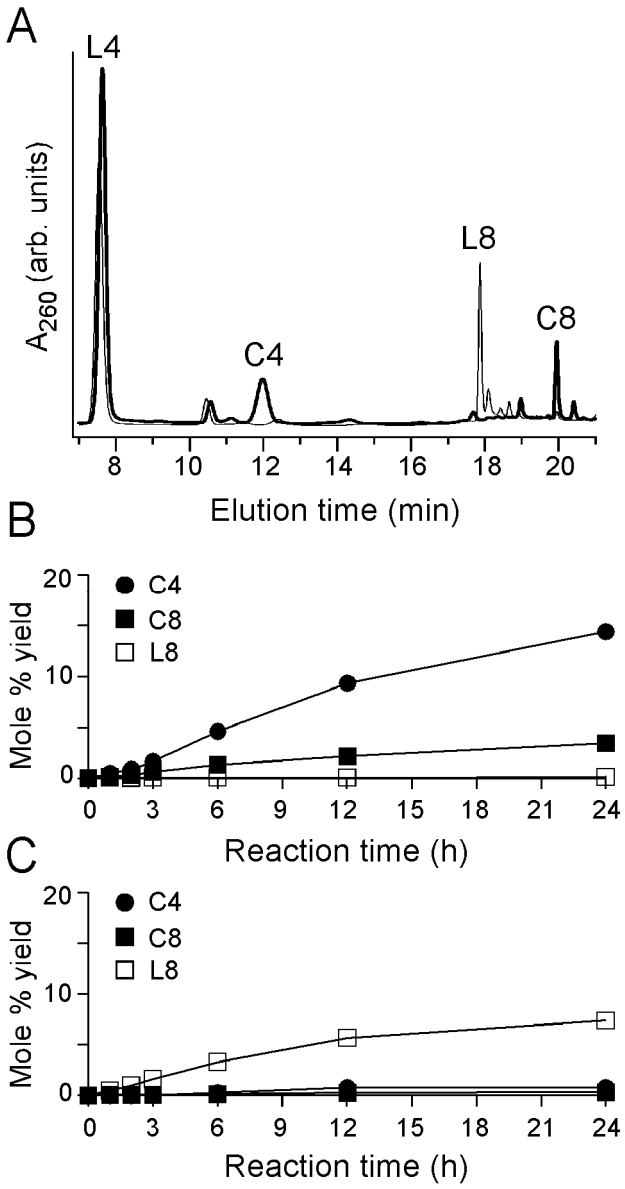


Figure 4.2. A) Representative HPLC chromatographs of products formed by d(pCGTA) (200 μM) with activation by the condensing agent *N*-cyanoimidazole (25 mM, 125 eq.), and incubation at 4 °C for 24 h. The *thick line* is from a reaction carried out in the absence of ethidium, and the *thin line* is from a reaction containing ethidium (600 μM, 3 eq.). Product labels are L4, linear tetranucleotide (starting material); C4, cyclic tetranucleotide; L8, linear octanucleotide; and C8, cyclic octanucleotide. B) Products formed by d(pCGTA) as a function of incubation time for the reaction described in *A* without an intercalator. C) Products formed by the same reaction shown in *B*, in the presence of the intercalator ethidium (600 μM, 3 eq). Reaction product yields were determined by separation using reverse-phase HPLC, integration of UV absorbance and normalization relative to d(pCGTA) starting material.

Phosphodiester bond formation is unfavorable without the aid of protein enzyme catalysis. Therefore previous demonstrations of non-enzymatic template directed synthesis have involved chemical activation of an oligonucleotide with a terminal phosphate. Cyanogen bromide, although effective, has a short lifetime in aqueous solution [63]. In contrast, the imidazole derivative N-cyanoimidazole is stable over the course of several days. N-cyanoimidazole activation produces an oligonucleotide with a corresponding phosphorimidazolide, which is a more reactive species than the native phosphate [63].

To test the ability of intercalators to promote oligonucleotide polymerization and prevent cyclization, d(pCGTA) was treated with the activating reagent N-cyanoimidazole, in the absence and presence of ethidium, a well-characterized intercalator of Watson–Crick duplexes. The early products of the ligation reaction (cyclic tetranucleotide, linear octanucleotide and cyclic octanucleotide) were monitored by HPLC. Example chromatograms and plots derived from the analysis of such chromatograms for reactions allowed to incubate for various times are shown in Figure 4.2. In the absence of ethidium, the major product is the cyclic tetranucleotide (Figure 4.2B), as should be expected due to the very short persistence length of single-stranded oligonucleotides. Only trace linear octanucleotide was formed, and about one-fourth as much cyclic octanucleotide as cyclic tetranucleotide was formed. The lack of appreciable linear octanucleotide formation indicates that octanucleotide cyclization occurs rapidly, compared to tetranucleotide ligation. In contrast, in the presence of ethidium, linear octanucleotide formation is greatly favored over tetranucleotide cyclization (Figure 4.2C). It was also observed that the rate of octanucleotide formation in the presence of

ethidium is slower than the tetranucleotide cyclization rate in the absence of intercalator. Thus, in addition to promoting linear product formation, ethidium inhibits strand cyclization, presumably by promoting assemblies with greater persistence lengths (Figure 4.1B).

4.3.2. Pre-ligation Assembly of Tetranucleotides

As stated previously by Hud and Anet [53], non-covalent assembly of nucleic acids is a necessary requirement for the formation of natural nucleic acids under prebiotic conditions. The tetranucleotide d(pCGTA) was heat denatured and monitored by UV-vis under the same conditions as the chemical ligation before the addition of N-cyanoimidazole. In the absence of ethidium no cooperative melting transition is observed (Figure 4.3). On the other hand, the addition of ethidium stabilizes the assembly. The tetranucleotide-ethidium assembly has a T_M = 30°C. Therefore at the chemical ligation reaction temperature of 4°C, the tetranucleotide is completely assembled into an intercalated duplex. ^{1}H-NMR shows that titration of ethidium into a solution of tetranucleotide creates a non-covalent molecular complex that is on the intermediate exchange time scale. The exchange regime is determined by the line broadening observed between the no ethidium and 200 μM ethidium spectra shown in Figure 4.4. At an ethidium concentration of 200 μM the aromatic ^{1}H-NMR signals have broadened line widths from *ca.* 5 Hz to 13 Hz and thereafter broaden into the baseline. The broad signals in the 600 μM ethidium (7.1 and 7.4 ppm) could be resonances coming from the excess ethidium present in the final titration solution.

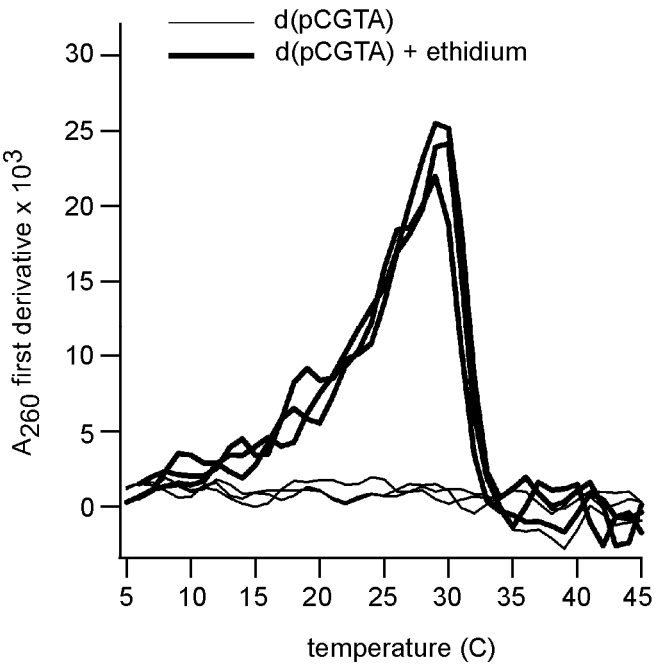


Figure 4.3. The first derivative curves of the UV-vis thermal denaturation of the tetranucleotide d(pCGTA) (200 μM) with and without ethidium (600 μM). The complexes shown are non-covalent as no N-cyanoimidazole was added. Sample heating and cooling curves are overlaid for both samples.

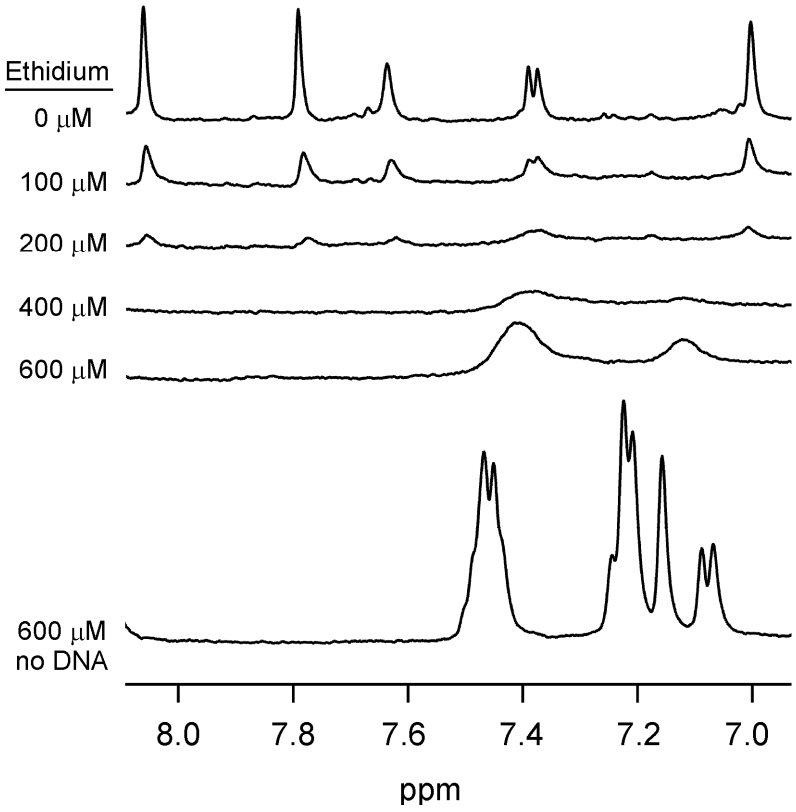


Figure 4.4. ^{1}H-NMR titration of ethidium into a solution of tetranucleotide (200 μM) containing 10 mM MES-TEA (pH 6) buffer and 5 mM $MgCl_2$. Magnesium is used rather than Manganese because of the paramagnetic properties of manganese. The complexes shown are non-covalent as no N-cyanoimidazole was added. The control sample with only ethidium and buffer is shown at the bottom.

4.3.3. Intercalators Promote Polymer Formation

Denaturing polyacrylamide gel electrophoresis (PAGE) and SYBR Gold staining was used to monitor the synthesis of longer ligation products in these reactions. The gel image shown in Figure 4.5 confirms that d(pCGTA) does not ligate appreciably in the absence of an intercalator. While the tetranucleotide starting material was not observable by SYBR Gold staining, the observation of small amounts of cyclic octanucleotide and cyclic dodecanucleotide is consistent with results obtained by HPLC. In contrast, when ethidium was added to the same reactions, polymers of up to 100 nucleotides in length

were observed (*i.e.,* 24 linear couplings) (Figure 4.5). This analysis also illustrates that the ratio of linear to cyclic products increases with increasing ethidium concentrations for all polymer lengths (Figure 4.5). For example, no linear octanucleotide or linear dodecanucleotide is detected in the absence of ethidium (lane 1, Figure 4.5), while approximately equal amounts of linear and cyclic octanucleotide and dodecanucleotide products are observed when ethidium is present at a stochiometry of one ethidium per tetranucleotide (lane 2, Figure 4.5). For higher ethidium to tetranucleotide stochiometeries, the relative amounts of cyclic products are far lower than linear products of the same nucleotide length (lanes 3-5, Figure 4.5).

The presence of ethidium does not prevent the formation of cyclic oligonucleotides altogether as evidenced by double bands in the lanes containing ethidium. Treatment of the chemical ligation reaction with Calf Intestine Alkaline Phosphatase (CIAP) removes the 5′ terminal phosphate on linear products whereas cyclic products remain unmodified. CIAP treatment of the tetranucleotide reaction containing 600 μM ethidium shows a decrease in the mobility of one of the bands corresponding to eight nucleotides while the other band remains unchanged (Figure 4.5 inset). This data shows that one of the bands is linear and the other is cyclic.

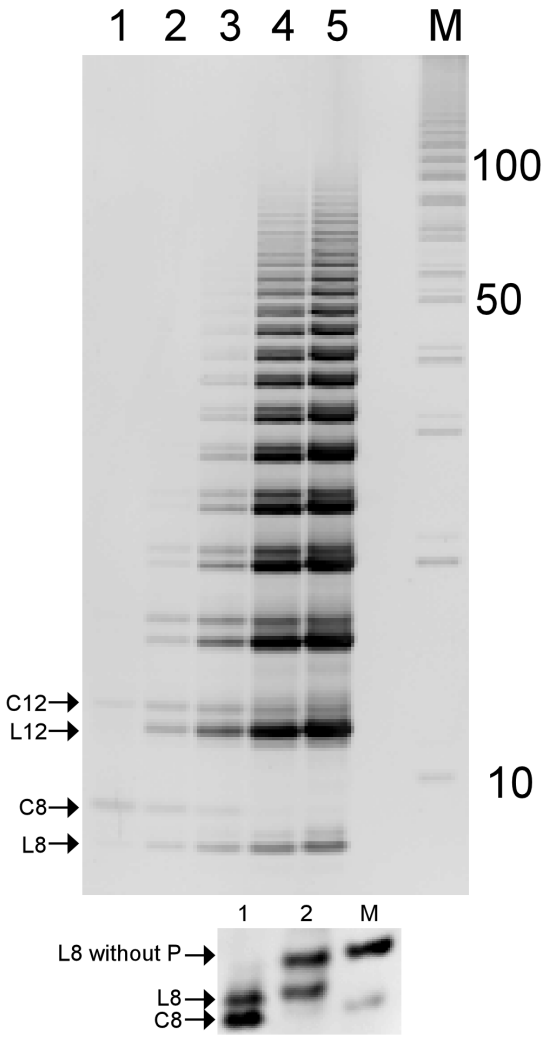


Figure 4.5. The addition of ethidium to the tetranucleotide d(pCGTA) prevents cyclization and promotes oligonucleotide polymerization. Lane 1: 200 μM d(pCGTA), condensed with 250 mM N-cyanoimidazole at 4ºC for 72 h. As indicated by arrows, with no intercalator present, only small amounts of cyclic octanucleotide (C8) and cyclic dodecanucleotide (C12) are produced and almost no linear octanucleotide (L8) or linear dodecanucleotide (L12). Lanes 2-5 illustrate that increasing ethidium:substate stoichiometries promote linear polymerization far beyond that observed in lane 1, to *ca.* 100 nucleotides in length. Ethidium concentrations for lanes 2-5 were 100 μM, 200 μM, 400 μM and 600 μM, respectively. The tetranucleotide starting material and cyclic tetranucleotide products observed in HPLC analyses were not observed here, due to inefficient staining. **Inset.** Lane1: 200 μM d(pCGTA), condensed with N-cyanoimidazole in the presence of 600 μM ethidium at 4ºC. Lane 2: Same as Lane 1 only after treatment with CIAP. Linear octanucleotide is labeled with and without a phosphate (L8 without P).

4.3.4. Ethidium-Mediated Ligation is Dependent on Watson–Crick Base Pairing

The tetranucleotide d(pCGTA) can form regularly nicked duplexes with Watson–Crick base pairs. In order to demonstrate that Watson–Crick base pairing is required for ethidium-mediated ligation (and ligation is not merely promoted by non-specific electrostatic or hydrophobic interactions), two other oligonucleotides, d(pCCTA) and d(pGGTA), were studied under the same ligation conditions as those used in the experiments presented above. Individually, neither of these two tetranucleotides is able to form a nicked duplex polymer with complete Watson–Crick base pairing, whereas a d(pCCTA)/d(pGGTA) mixture can form a fully-paired and regularly nicked duplex. PAGE analysis confirmed that, following condensation activation by N-cyanoimidazole, neither tetranucleotide alone forms appreciable ligation products in the absence or presence of ethidium (Figure 4.6). In contrast, efficient polymerization occurs when the mixture of the two tetranucleotides is activated in the presence of ethidium (Figure 4.6), confirming the importance of Watson–Crick base pairing in ethidium-mediated ligation.

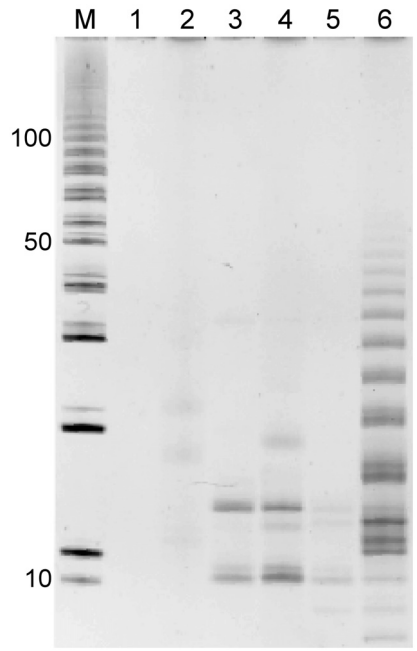


Figure 4.6. Condensation reactions demonstrating that ethidium-mediated oligonucleotide polymerization requires Watson–Crick base pairs. Lanes 1 and 2: d(pCCTA) (200 μM) with and without ethidium (600 μM), respectively. Lanes 3 and 4: d(pGGTA) (200 μM) with and without ethidium (600 μM), respectively. Lanes 5 and 6: d(pCCTA)/d(pGGTA) mixture (100 μM each) with and without ethidium (600 μM), respectively. All reactions were carried out at 4ºC for 72 h with 250 mM N-cyanoimidazole. The efficient ligation demonstrated in Lane 6 was only observed with the combination of fully complementary oligonucleotides in the presence of ethidium. The bands observed in Lanes 3 and 4 apparently result from the coupling of oligonucleotides paired by non-Watson–Crick base pairs (*e.g.,* G·G mismatches), as these side products are not enhanced by the addition of ethidium.

4.3.5. Base Pair Recognition is Important (but not sufficient) for Intercalator-Mediated Ligation

Our investigations of a variety of planar molecules as potential midwife molecules provide support for our proposal that recognition of a particular base pairing motif is a necessary criterion for intercalator-mediated ligation. Tetranucleotide condensation experiments with coralyne illustrate this point. Coralyne is a planar molecule that is somewhat larger than ethidium (Figure 4.7), but also has a single positive

charge. Coralyne binds to Watson–Crick DNA with an association constant that is within *ca.* a factor of two of the association constant of ethidium for Watson–Crick DNA [57]. Nevertheless, condensation experiments demonstrate that coralyne does promote polymerization of the tetranucleotide d(pCGTA) (Figure 4.7). Base pair recognition alone by an intercalator is, however, not sufficient for achieving intercalation-mediated polymerization. For example, proflavine (an intercalator that binds duplex DNA with similar affinity to ethidium) does not promote ligation in our Watson–Crick pairing system (Figure 4.7).

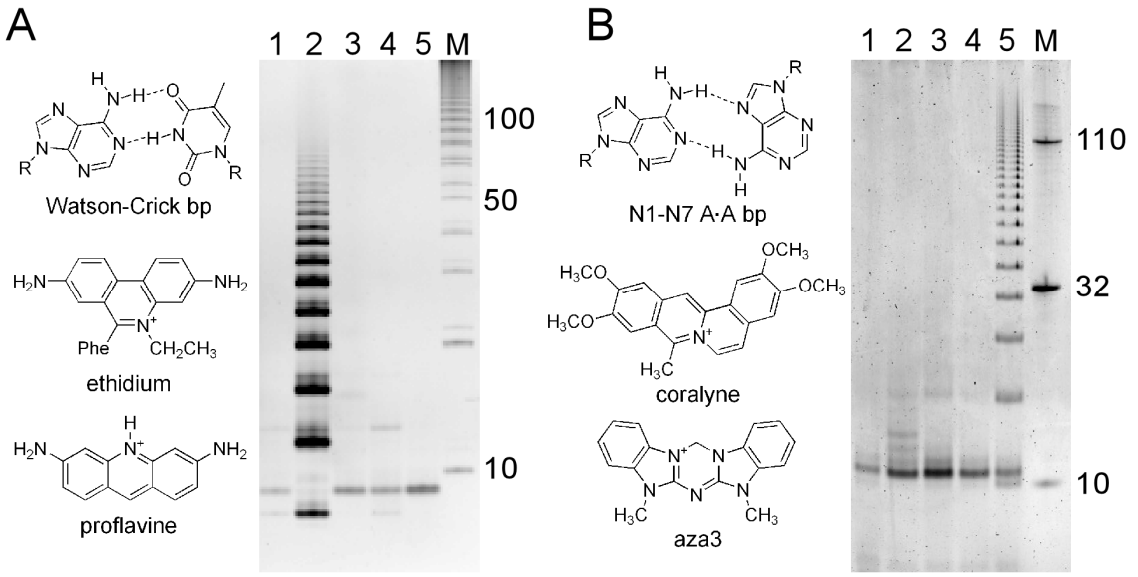


Figure 4.7. Base pair recognition by a given intercalator is an important, but not exclusive, criterion for achieving intercalation mediated ligation. A) PAGE analysis of products from the condensation of d(pCGTA) (200 μM) in the absence and the presence of various small molecules (600 μM). Ethidium promotes the ligation of d(pCGTA) into long polymers (lane 2). In contrast, proflavine, another Watson–Crick intercalator does not promote the ligation of d(pCGTA). Neither coralyne (lane 4) or aza3 (lane 5), two intercalators that stabilize A·A base pairs, promotes d(pCGTA) ligation beyond that observed in the absence of an intercalator (lane 1). B) PAGE analysis of products formed by the condensation of d(pA_6) (500 μM) in the absence and the presence of the same small molecules (750 μM) as A. Neither Watson–Crick intercalator promotes the ligation of d(pA_6) (lanes 2 and 3) beyond that observed in the absence of intercalator (lane 1). In contrast aza3 promotes the ligation of d(pA_6) (lane 5). Although coralyne also stabilizes A·A base pairs it does not promote the ligation of d(pA_6) (lane 4). The N1-N7 base pair shown for homo-adenine duplexes has been shown to be consistent with molecular modeling and chemical probing studies (Joung, et al., submitted). For more experimental details see Materials and Methods.

4.3.6. Intercalation-Mediated Ligation with non-Watson–Crick Base Pairs

Intercalator-mediated polymerization should be possible with a variety of base pairing structures. As one example, we have previously demonstrated that an intercalator which binds pyrimidine triplexes (*i.e.,* triple helical nucleic acids formed by a Watson–Crick duplex and a third homo-pyrimidine strand) can be used to promote the ligation of the third strand on a Watson–Crick template [54]. In principle, it should also be possible

drive the polymerization of oligonucleotides that are held together with non-Watson–Crick base pairs that are unstable in the absence of ligand binding, given an intercalator that stabilizes the alternative base pairs. Our discovery that the crescent-shaped molecule coralyne and aza3 (Figure 4.7) stabilize duplex nucleic acids with A·A base pairs [59,60,142] provided a means to test this proposal. In this set of experiments, the hexanucleotide d(pA_6) was used as the starting oligonucleotide, rather than a tetranucleotide, as the aza3-stabilized A·A duplex is less stable than an ethidium-stabilized Watson–Crick duplex. PAGE analysis demonstrates that aza3 efficiently promotes the ligation of d(pA_6) upon activation with N-cyanoimidazole. In contrast, ethidium, shown above to promote ligation of oligonucleotides that form Watson–Crick base pairs, does not promote ligation of d(pA_6) (Figure 4.7). Similarly, aza3, which has low affinity for duplexes with Watson–Crick base pairs [165], does not promote the ligation of d(pCGTA) (Figure 4.7). We note that although coralyne also stabilizes duplexes with A·A base pairs, coralyne does not show evidence of promoting d(pA_6) ligation. Together, these data emphasize the requirement of recognition between an intercalator and a particular base pair for promoting intercalation-mediated polymerization, and again illustrate that base pair recognition alone is not sufficient for promoting intercalation-mediated polymerization.

4.4. DISCUSSION

It has long been postulated that protein enzyme-free synthesis and template-directed replication of nucleic acid polymers (or their predecessors) was an early and critical step in the origin of life [166-169]. Polymer cyclization is an inherent obstacle to

the growth of polymers from bifunctional mononucleotides and oligonucleotides, and it is one that would have potentially thwarted the abiotic synthesis of the first RNA-like polymers, unless a mechanism existed by which mononucleotides and short oligonucleotides were organized prior to chemical ligation. Previous examples of non-enzymatic template-directed polymerization have required relatively high concentrations of substrates [159] or chemical protection of one end, in the case of oligonucleotide ligation, in order to prevent strand cyclization [167]. The experiments presented here demonstrate that strand cyclization can be circumvented by molecules that intercalate the base pairs of nucleic acid duplexes. It is possible that the abiotic reactions which originally gave rise to the nucleotide bases would have also produced a wider variety of appropriately sized heterocycles (*e.g.,* flavins [170]), that might have functioned as molecular midwives for the assembly of the first RNA-like polymers. Furthermore, intercalation provides additional free energy to duplex stability, allowing ligation reactions to occur at concentrations that are orders of magnitude lower than those which have been previously utilized in model prebiotic systems.

In the present study, we have presented results from experiments involving the ligation of DNA oligonucleotides, which provide a convenient and robust model system for exploring intercalation-mediated polymerization. These results are, nevertheless, also directly relevant to theories regarding the non-enzymatic polymerization of RNA, as similar experiments with RNA oligonucleotides also show enhanced ligation in the presence of ethidium, albeit to a lesser extent than DNA (data not shown). The diminished enhancement of RNA ligation may be due to the fact that the intercalators discussed in the present work bind RNA with a lower affinity as compared to DNA.

Additionally, the product bands observed upon PAGE analysis of RNA ligation reactions were more diffuse, possibly because RNA can form both 3′,5′ and 2′,5′ linkages, whereas DNA cannot form the 2′,5′ linkage. This lack of regioselectivity in chemical ligations of RNA has been noted previously [64].

The chemical ligation studies presented here have revealed the sensitivity of our model ligation system to the structure of the intercalator used to promote assembly. In particular, proflavine did not show activity in promoting ligation, even though proflavine and ethidium have similar association constants for duplex DNA. Interestingly, in a previous study of a ligation system using a different activation chemistry, we found that proflavine, but not ethidium, promotes oligonucleotide ligation [61]. Thus, the efficiency of an intercalator for promoting a given ligation reaction appears not to be fully expressed by its association constant for a nucleic acid assembly. In retrospect, this observation is not surprising; covalent bond formation requires an accessible reactant conformer with the appropriate geometry, and individual intercalators make structure-specific helix contacts, which would be expected to modify the conformational landscape of the polymer backbone in an intercalator-specific manner. This feature could prove useful for selecting specific backbone linkages in intercalation-mediated reactions.

Here, we have used tetranucleotides as a model system to illustrate the enzyme-free production of polymers from short oligonucleotides. With regards to a more complete model for the origin of RNA-like polymers, it is certainly desirable to demonstrate that polymerization can be achieved with even shorter oligonucleotides and, ultimately, mononucleotides. Thus far, our attempts to use ethidium to drive the polymerization of mononucleotides and dinucleotides in the absence of a pre-existing

template strand have proven unsuccessful. In the case of dinucleotides, we surmise that the nearest neighbor exclusion principle [45], which states that intercalators can bind at most between every other base pair, is potentially inhibiting the coupling of dinucleotides, as stacked dinucleotides that are assembled by intercalators tend to have intercalators bound in between each dinucleotide [32], an arrangement that could block backbone coupling between dinucleotides. To find a possible means to circumvent this problem, we have initiated studies of intercalation of nucleic acids with alternative backbone linkages [73,103]. Given the wide variety of RNA-like polymers with sugars other than ribose that also form duplexes [68,171,172]; it seems possible that the backbone of the first informational polymers may not have utilized a ribose sugar. This original pre-RNA backbone might have allowed intercalation between every base pair. With regards to mononucleotide polymerization, it appears that the poor stacking of the pyrimidine nucleotides precludes assembly of mononucleotides by intercalating molecular midwives. The poor stacking of pyrimidine bases has prompted our laboratory, and several others [173-175], to reconsider the proposal that the first RNA-like polymers used purine-purine base pairs [176]. Our demonstration in the present work that intercalators can promote the polymerization of homo-adenine oligonucleotides represents an early step towards the realization of intercalation-mediated mononucleotide polymerization in a homo-purine system.

The experimental observation of broad bands in the ligation of d(pCCTA) and d(pGGTA) illustrates the point that the intercalation mediated-assembly of non-identical tetranucleotides may result in the creation of quasi-random sequence information (new base sequences which did not originate in the initial solution), albeit much less

information than is possible if all nucleotide positions are allowed to vary. In a pool of random oligonucleotides, each tetranucleotide would have its complement present at a ratio of 1 complement per 256 sequences. In contrast, sequences with two-base overhangs would be present at a ratio of 1 per 16 sequences. Investigations in the Hud lab are currently directed towards demonstrating that, a random pool of tetranucleotides has a high probability of forming nicked duplexes. Stabilization of such random, continuously-nicked duplexes with intercalators would provide a convenient route to nucleic acid polymers with sequence diversities that increase exponentially with length.

Recently, Szostak and coworkers demonstrated the non-enzymatic template-directed synthesis of RNA inside model protocells [177]. These model protocells exhibited high permeability for small, minimally charged species. Therefore, molecules similar to the intercalator midwives discussed in the present work, which contain only a single positive charge, could have permeated protocells along with nucleotides, before the advent of protein enzymes, and assisted in the assembly of encapsulated nucleic acid polymers.

Finally, it was recently proposed that the tendency for Watson–Crick DNA duplexes to form liquid crystals could reflect the process by which prebiotic nucleic acids were organized (and partitioned) prior to polymerization [178]. The experimental results presented here demonstrate that base pair intercalators can promote the assembly of oligonucleotides that are shorter and at concentrations far lower than those required for liquid crystal formation. Additionally, the selectivity of a given intercalator for a specific base pair (*e.g.,* Watson–Crick base pairs) suggests a mechanism by which the earliest base pairings of life might have been selected.

4.5. CONCLUSION

Polymer cyclization is a difficulty inherent to the growth of polymers from bifunctional monomers and oligonucleotides, and is one that would have potentially thwarted the abiotic synthesis of the first RNA-like polymers unless a mechanism existed by which short oligonucleotides were organized prior to chemical ligation. The experimental results presented here demonstrate that small, heterocyclic ligands that intercalate Watson–Crick base pairs, which could have been created by the same chemistry that produced the nucleotide bases [179,180], can promote the assembly of oligonucleotides that are shorter and at concentrations orders of magnitude lower than those required for liquid crystal formation. Additionally, the selectivity of a given intercalator for a specific base step size (*e.g.,* duplexes with Watson–Crick base pairs) [142] suggests a mechanism by which the first base pair structures might have been selected. Furthermore, using the d(pCCTA)·d(pGGTA) system we have created a semi-random pool of oligonucleotides of which deoxyribozymes could be selected from.

The observation that an ethidium concentration that is higher than what is expected for nearest-neighbor loading and the observation of some cyclic products even in the presence of ethidium suggests an intercalator with a higher affinity could therefore prevent cyclization further and also create longer polymers. It would also be required for this intercalator to be able to place the 3′ OH and the activated 5′ phosphate into a ligation-accessible state. The Hud laboratory is in constant pursuit of ‘simple’ intercalators which would bind Watson–Crick base pairs with high affinity.

Although the need for tighter-binding intercalators as well as the requirement for structural activity in the ligation complex appears to be a major drawback, consideration

must be placed to the idea of non-natural nucleic acids as well as different forms of ligation chemistry such as backbone linkages that are thermodynamically driven [66].

CHAPTER 5

FUTURE DIRECTIONS AND CONCLUDING REMARKS

5.1. FUTURE DIRECTIONS

5.1.1. Intercalator-Mediated Stabilization of Hybrid Duplexes

In some presentations of the RNA world hypothesis it is suggested that the precursors of RNA, *e.g.,* bases, ribose and phosphate, were formed spontaneously on the prebiotic Earth, before the emergence of what might be considered a metabolism. For a number of reasons, it is difficult to accept that a pool of precursors, much less a reservoir of chiral pure nucleotides, existed as a mixture of only those precursors capable of forming RNA. For example, the formose reaction has been cited as possible source of prebiotic sugars [181,182]. However, ribose is only a minor product amongst the large variety sugars produced by this reaction. Thus, without a clear mechanism by which ribose would have been selected out of a pool of diverse sugars, it is considered by some researches more plausible that nucleic acids were originally formed with a different sugar in place of ribose, or even a heterogeneous assortment of sugars [53].

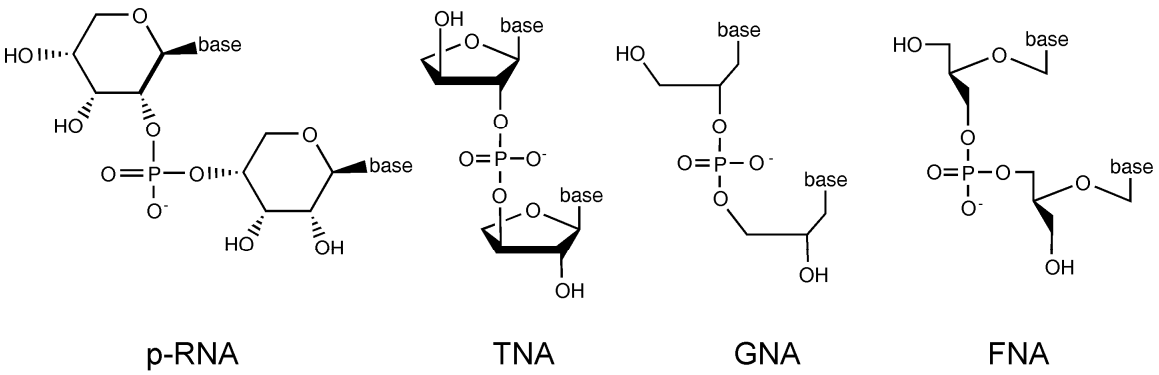


Figure 5.1. Schematic of the nucleic acid analogs pyranosyl RNA (p-RNA), threose nucleic acid (TNA), glycerol nucleic acid (GNA), and flexible nucleic acid (FNA).

The Eschenmoser laboratory has demonstrated that a number of nucleic acids with bases other than ribose are able to form stable duplexes, as well as cross-pair with RNA (*i.e.,* form hetero-duplexes) [68]. The ability for cross-pairing between nucleic acids with different sugars has been considered by several origin of life researchers to be a strict test for the feasibility for the transfer of genetic information from a pre-RNA world to the RNA world [183]. It was shown by Eschenmoser and coworkers that the family of pentopyranosyl nucleic acids (Figure 5.1) linked between the 4′ and 2′ hydroxyls are capable of cross-pairing with the natural pentofuranosyl form of RNA [184]. In addition, threose nucleic acid (TNA), which utilizes a shorter backbone, but is still in the furanosyl conformation, can also cross-pair with RNA [171]. Finally, even nucleic acids with acyclic sugars, such as the acyclic TNA derivative GNA (glycerol nucleic acid), can cross pair with RNA [185].

On the other hand, there are also a number of nucleic acid backbones that have been synthesized which do not cross-pair, *i.e.,* form stable duplexes, with natural RNA. One example is the flexible nucleic acid backbone (FNA) (Figure 5.1), which was shown by Merle *et al.* not to form stable homo-duplexes or hetero duplexes with natural nucleic

acids. Before it was demonstrated that FNA cannot cross-pair with RNA, this nucleic acid was suggested as a likely precursor to RNA because of its relatively simple, acyclic chemical structure [186,187]. Furthermore, the glycerol unit of FNA is stereochemically analogous to a ribose moiety in which the C2′ carbon has been removed (or not yet inserted). More recently, Switzer and coworkers have demonstrated that the DNA polymerase *Therminator* is able to polymerize FNA-triphosphates using a DNA template [146]. While the DNA polymerase is not necessarily prebiotic, this study did prove that hybrid systems that do not form stable duplexes are, in some cases, able to transfer sequence information from a template to a product strand.

2′,5′ RNA, the structural isomer of natural RNA discussed in Chapters 2 and 3, forms stable hybrid duplexes with natural RNA [100]. In contrast, Damha and coworkers observed that 2′,5′ RNA does not form stable duplexes with 3′,5′ DNA [100,188]. The high association constant and backbone interactions demonstrated in previous Chapters for 2′,5′ RNA binding to proflavine provides an excellent system to investigate the ability for intercalators to stabilize hybrid duplexes that are not stable in the absence of intercalators.

The addition of single stranded 2′,5′ RNA and single stranded 3′,5′ DNA produce a CD spectrum that is not a linear combination of the single strand CD spectra (Figure 5.2). This result indicates some level of association between the two oligonucleotide strands, which is contrary to the conclusion of Damha and coworkers that 2′,5′ RNA and 3′,5′ DNA do not hybridize [100]. The oligonucleotides used in these experiments contained only G and C residues to provide maximum stability, and nucleotide sequences were specifically designed to have minimal self-structure (*e.g.,* hairpins, self-dimers,

quadruplexes). Therefore, the observation of a 2′,5′ RNA/3′,5′ RNA hybrid duplex may be due to the added stability of having all G·C pairs. Samples containing either of the two oligonucleotide strands do not exhibit a cooperative melting transition when monitored by CD.

While the cooperative transition observed for the 2′,5′ RNA/3′,5′ RNA hybrid duplex without proflavine is *ca.* 35°C, the addition of proflavine creates an ICD in the visible region between 400 and 500 nm (Figure 5.2) (similar to the ICD observed in Chapter 3) as well as increases the T_M by *ca.* 15°C. The addition of proflavine apparently promotes the formation of a self-structure in samples containing either of the single strands, as evidenced by the ICD bands. The linear combination of these control spectra are very similar to the spectrum produced by the hybrid duplex in the presence of proflavine. Thermal denaturation of the single strands with proflavine does not display cooperative melting (Figure 5.3).

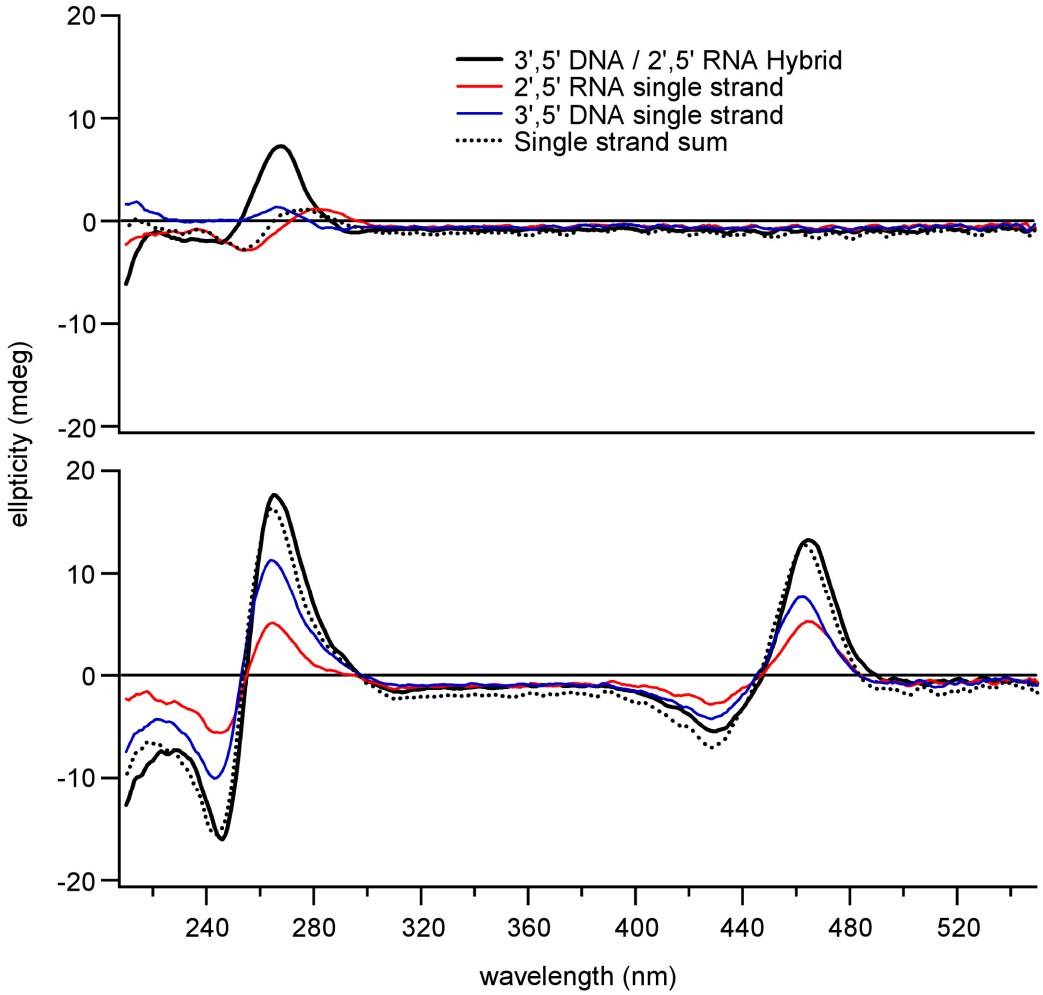


Figure 5.2. CD spectra of the 2′,5′ RNA/3′,5′ DNA hybrid duplex without (Top) and with (Bottom) proflavine. Strand sequences were 3′,5′ DNA: 5′-CCCGCCGCGCCG-3′, 2′,5′ RNA: 5′-CGGCGCGGCGGG-2′. Oligonucleotide concentrations were 40 μM bp for hybrid spectra and 40 μM base for single stranded spectra. Samples containing intercalator had 20 μM proflavine. CD spectra were acquired at 5°C in 1× BPE, 100 mM NaCl buffer.

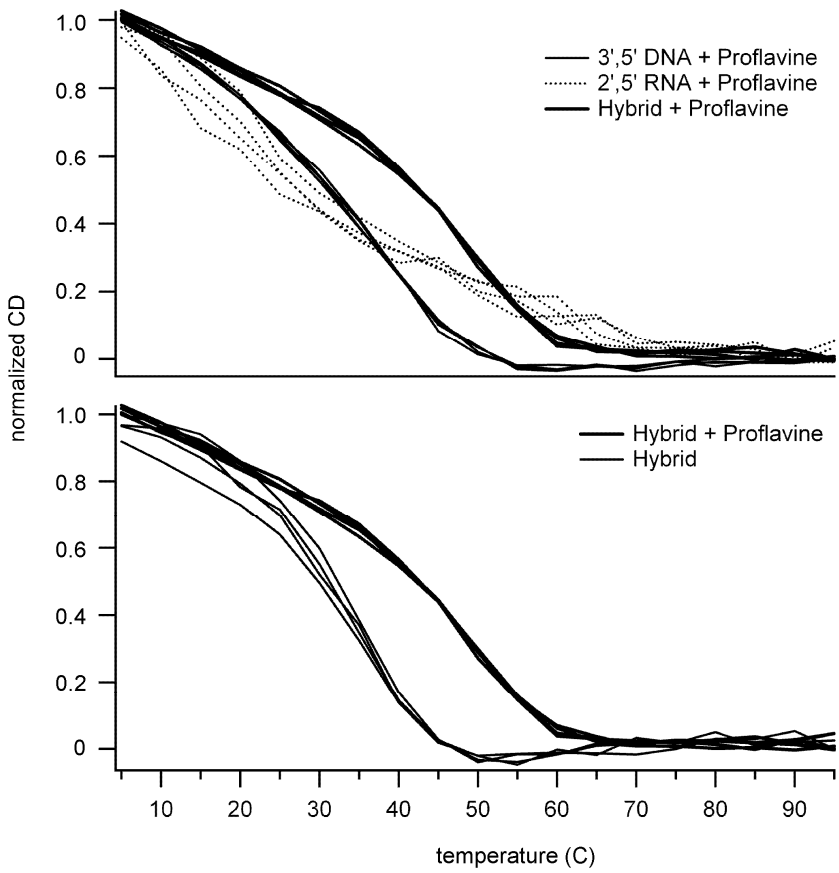


Figure 5.3. Top: Normalized CD thermal denaturation of the 3′,5′ DNA and 2′,5′ RNA single strands and the hybrid duplex with proflavine. Bottom: Normalized CD thermal denaturation of a 2′,5′ RNA/3′,5′ DNA hybrid with and without proflavine. Melting curves were normalized by fitting each spectrum to a linear combination of the first (5°C) and last spectra (95°C) in the UV region. Pre-transition slopes indicate duplex destacking prior to the cooperative melt. Multiple traces represent two heating and two cooling cycles.

The observed increase in hybrid duplex stability is of similar magnitude to that observed for the addition of proflavine to the 2′,5′ RNA/2′,5′ RNA homo-duplex, as discussed in Chapter 2. While this datum is preliminary, and does not prove that the association of proflavine with the hybrid duplex is intercalation, the fact that proflavine intercalates 3′,5′ DNA as well as 2′,5′ RNA homo-duplexes (see Chapter 3), strongly suggests that proflavine can also intercalate the hybrid duplex. In any case, these

experiments demonstrate the potential for an intercalator to promote duplex formation between oligonucleotides with different backbones that do not otherwise hybridize.

5.1.2. Predisposing Base Steps for Intercalation

As discussed in Chapter 3, an alternating sugar pucker arrangement at an intercalation site (C3′ *endo*/C2′ *endo* in DNA and RNA, also described as compact/extended inter-phosphate distances) is energetically favorable [49]. In Chapter 3 it was also shown that when the RNA linkage is changed from 3′,5′ to 2′,5′, a preference of alternating sugar puckers is observed, which is C2′ *endo*/C3′ *endo*, but this pattern is again of compact/extended inter-phosphate distances in the case of 2′,5′ RNA.

Although there are examples of DNA intercalation where the alternating sugar pucker arrangement is not observed [47,97], our experimental results as well as computational studies indicate [49], that the alternating sugar pucker state is the most energetically favored state of the nucleic acid backbone at a site of intercalation. Chaires has shown that the amount of energy required to create an intercalation site is on the order of +5-10 kcal mol^{-1} [38]. It is therefore reasonable to predict that the preorganization of an oligonucleotide with these alternating compact/extended sugar conformations would decrease the amount of energy required to form an intercalation site.

There is more than one way by which it may be possible to preorganize a nucleic acid backbone for preferential site intercalation. Locked-nucleic acids (or LNA) consist of nucleotides with a methylene bridge between O2′ and C4′ positions of a ribose nucleoside (Figure 5.4), which “locks” the nucleotide into the C3′ *endo* sugar pucker

conformation. It has been demonstrated that these modifications increase the stability of an LNA-RNA hybrid duplex because no energy is required to switch the LNA nucleotide from the preferred DNA sugar pucker of C2′-*endo* (an extended or south pucker conformation) to the preferred C3′-*endo* sugar pucker (a compact or north pucker conformation) of a hybrid duplex formed with RNA [189]. In 2007 Jayakumar *et al*. also reported that incorporation of a 3′-phosphorothioate induces the attached sugar convert to a C3′ *endo* sugar pucker conformation [190]. For example, incorporation of a phosphorothioate (p**s**) at a CpG step (*e.g., C*p**s***G*) causes the C sugar to be in the C3′ *endo* conformation, while the G would remain in the C2′ *endo* conformation, which is the preferred sugar pucker arrangement for intercalation.

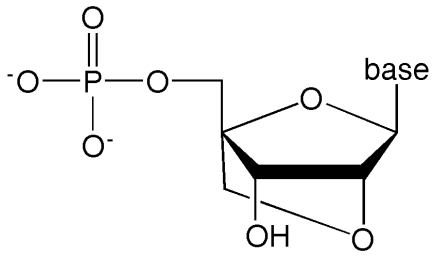


Figure 5.4. A locked nucleic acid mononucleotide (LNA). The O2′ is bridged to the C4′ through a methylene bridge.

Although incorporation of LNA (**l**) nucleotides (*e.g.,* **l**CpG) could cause the same effect, the addition of the methylene bridge to the nucleotide may provide steric hindrance, which may decrease the association constant of the intercalator. To test the hypothesis that pre-disposing the nucleic acid backbone into alternating C3′ *endo*/C2′ *endo* steps makes base pair intercalation more energetically favorable, the effect of both modifications will be determined. Oligonucleotides will require at least one internal CpG step because of the observed preference for base pair intercalation at CpG steps. For

example the self-complementary dodecanucleotide d(GpGpCpGpApTpCpGpCpCp) has two CpG steps as well as two C·G base pairs that flank the CpG steps. The central hex nucleotide of this sequence has been used previously for crystallographic as well as NMR studies of nucleic acid intercalation. The CpG steps will be modified with either phosphorothioate linkages or LNA sugars to predispose the intercalation site (d(GpGpCp**s**GpApTpCp**s**GpCpCp) and d(GpGp)**l**(Cp)d(GpApTp)**l**(Cp)d(GpCpCp) respectively). Although prediction of binding site preference is conceptually straight forward, the problems borne out by experiment may produce unwanted results. To prove that the intercalator is binding at the two CpG steps either NMR or crystallography must be employed to determine the sugar puckers before intercalation and to determine the binding sites of the small molecules used.

5.1.3. Intercalation-Mediated Ligation: Generating Sequence Diversity

In vitro evolution of RNA has led to ribozymes capable of performing a variety of catalytic functions that are assumed to have been necessary for a basic metabolism and gene replication in a pre-protein RNA world. For example, Lincoln and Joyce have recently demonstrated the evolution of ribozyme that is capable of catalyzing its own formation from a pool of substrates that constitute two parts of the complete ribozyme [12]. Despite this impressive result, and the support it lends to the RNA world hypothesis, without an initial pool of nucleic acids with sufficient sequence diversity, the appearance of such ribozymes would not have been possible in the pre-RNA world.

Hud and Anet proposed that molecular midwives could be used to assemble short nucleic acid duplexes with high sequence diversity because of the lack of sequence

specificity inherent to intercalator-base pair stacking [53]. The data shown in Chapter 4, specifically the ligation of d(pCCTA)·d(pGGTA) assembled with ethidium, is akin to this in the possibility of creating longer polymers with sequence diversity that increases exponentially with length. By merely increasing the number of original tetranucleotides in the complex from one to two the number of potential products increases tremendously. To further prove this point it would be prudent to incorporate tetranucleotides with more randomized positions as well as demonstrate the ligation of shorter oligonucleotides and mononucleotides.

5.1.4. Intercalator/Ligation Chemistry Specificity

Swapan S. Jain, a previous member of the Hud laboratory, demonstrated that template-directed ligation of d(T_3-p**s**) and d(5′-iodo-T_4) was greatly enhanced in the presence of proflavine (i.e. about 1000-fold), but not in the presence of ethidium [61]. In contrast, the experiments presented in Chapter 4 demonstrate that N-cyanoimidazole phosphate activation of the oligonucleotide d(pCGTA) results in efficient polymer formation in the presence of ethidium, but not in the presence of proflavine. Additionally, the small molecule aza3, which binds duplexes with A·A base pairs, promotes the formation of linear polymers from d(pA_6), whereas coralyne does not, even though coralyne binds more tightly to A·A base pairs [142]. These observations emphasize that the ability to intercalate and stabilize a duplex with specific base pairs is not sufficient criteria to predict that an intercalator will create a ligation-active complex for a given backbone coupling chemistry.

These ligation systems were designed to create only one product (*i.e.,* 3′,5′-linked DNA). But it was previously demonstrated that given a choice of linkage (*i.e.,* 3′,5′ or 2′,5′ linkage), the product is highly dependent on the solution conditions [169] as well as the template used [191]. Clearly the pre-ligation state of the nucleic acid is critical to which backbone regio-isomer is formed.

As shown in Chapter 3, proflavine has H-bonds to the phosphate backbone when it intercalates 2′,5′ RNA. It would therefore be interesting to observe the intercalator-assisted ligation of 2′,5′ RNA polymers. The molecular midwife hypothesis [53] discusses recognition of the sizes of the base pairs by having optimal overlap during π-π stacking of the intercalator and the two base pairs of its binding site. Our observation of distinct thermodynamics for intercalation of 2′,5′ versus 3′,5′ RNA suggests that intercalator backbone interactions could also influence the backbone selected during ligation. Intercalation of a 2′,5′ RNA duplex prior to ligation may place the backbone in a conformation that directs a single backbone regio-isomer.

5.2. CONCLUDING REMARKS

Nucleic acid intercalation has been studied extensively for over four decades [72]. Although the structure and energetics of DNA and RNA intercalation have been discussed in great detail, there are still some fundamental questions that remain unanswered. For example, several explanations have been presented for the origin of the nearest-neighbor exclusion principle [45], but there is presently no consensus regarding the thermodynamic or structural origin of this principle [46,47,49,52,97]. It stands to reason that a more detailed understanding of intercalation thermodynamics, including the

contributions from the backbone, should help ultimately reveal the origin of the nearest-neighbor exclusion principle, as well as provide additional insights regarding the general nature of small molecule binding to nucleic acids by intercalation.

The structure and dynamics of the nucleic acid backbone should be expected to contribute significantly to the thermodynamics of intercalation. It has been shown in this thesis that investigating the effects of structural changes in the backbone on the energetic and structural transitions associated with intercalation can be very fruitful. In particular, the work presented in Chapters 2 and 3 demonstrate that intercalation of 2′,5′ RNA results in structural transitions that are similar to those observed for natural RNA. For example, as expected, the intercalation binding site is associated with helix unwinding and base pair destacking of the 2′,5′ RNA duplex . The increase in helical rise is to nearly two times the thickness of a π-stack, which is similar to that observed upon intercalation of natural RNA by proflavine [47]. Furthermore, both nucleic acid backbones exhibit an intercalator binding site preference for CpG steps.

A detailed structural investigation has shown that the formation of an intercalation site in 2′,5′ RNA also involves the unwinding of the phosphate dihedral angle ζ (similar as to what was observed for natural nucleic acids [125]). At the same time, 2′,5′ RNA exhibits another structural transition at the 2′ end of the intercalation site (guanosine). The sugar pucker of this residue changes from that of the unintercalated duplex, C2′ *endo* [87], to C3′ *endo*, which produces the alternating C2′ *endo*/C3′ *endo* conformation at the intercalation site. Although these sugar puckers are the opposite of what is commonly observed upon intercalation of natural RNA (C3′ *endo*/C2′ *endo*), the inter-phosphate distances are the same (compact /extended for the pyrimidine and purine respectively).

Therefore, the work presented in this thesis supports the hypothesis discussed previously by Voet [46] that the alternating sugar pucker motif is energetically favored in intercalation. The studies discussed in Chapters 2 and 3 represent what could be considered the first example of exploring the vast structural space of possible nucleic acid backbones that allow intercalation, perhaps providing more insight into the origins of some of the characteristics of intercalation and potentially revealing backbones that are even more amenable to intercalation (*e.g.*, backbones that allow violation of the nearest-neighbor exclusion principle).

With regards to the origin of nucleic acids, intercalating molecules may provide solutions to a number of problems associated with the search for how the first RNA-like molecules could have formed without the aid of protein enzymes. Along these lines, the studies discussed in this thesis are supportive of the ideas outlined in the molecular midwife hypothesis [53]. In this hypothesis the nucleic acid base pairs are proposed to have pre-assembled into columnar stacks with alternating base pairs and intercalator-like midwife molecules. The nucleic acid bases in these stacks were then covalently linked by a suitable backbone.

Although there are a variety of intercalator-like molecules that may have been present on the prebiotic Earth [151,179], molecules that would have acted as the proposed midwife molecules would have interacted with the bases in a manner that was compatible with the formation of an available backbone linkage (Chapters 2 and 3). From the work presented in this thesis, it is clear that not all intercalating molecules are compatible with all ligation chemistries. As shown in Chapter 4, in the presence of ethidium polymers up to 100 nucleotides can be formed from N-cyanoimidazole-

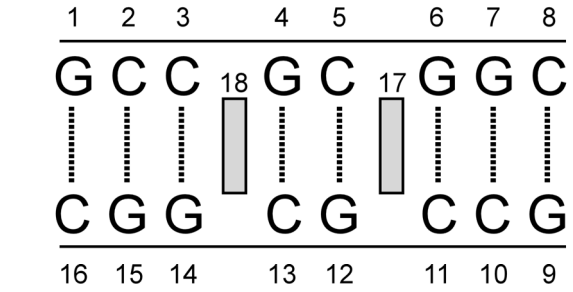
activated tetranucleotides. In contrast, proflavine, which has a similar association constant to DNA as ethidium [73], does not promote linear ligation of the same activated tetranucleotides. Similarly, with the d(pA_6) experiments shown in Chapter 4, there is a ligation preference for aza3 over coralyne, even though both have been shown to assemble homo-A duplexes [60,165].

Even for a midwife molecule that is compatible with a particular ligation chemistry, it might still exhibit a selective pressure on the exact structure of the backbone formed. For a chemistry that is compatible with proflavine acting as a midwife molecule, backbone formation in the presence of proflavine could be biased to 2′,5′ RNA linkages rather than 3′,5′ RNA linkages, as it has been shown in this thesis that proflavine binds 2′,5′ RNA with higher affinity.

All together, the work presented in this thesis is consistent with the molecular midwife hypothesis [53] and provides some glimpse to how intercalating molecules could have helped nucleic acids circumvent some of the problems that would have otherwise thwarted polymer growth (*i.e.,* strand cyclization, incompatible hetero-duplexes, backbone diversity). Although these data do not prove the molecular midwife hypothesis, the number of problems that can be solved by this relatively simple solution suggests that intercalators may have made significant contributions to the origin of RNA-like molecules on the prebiotic Earth.

APPENDIX A

SOLUTION STRUCTURE CONSTRAINTS



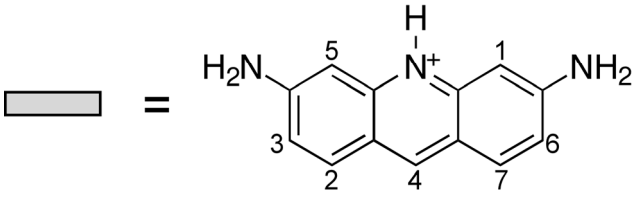


Figure A.1. Residue numbering schematic for the AMBER model. Terminal guanines are RGX residues while internal guanines are RGY residues. Terminal cytosines are RC3 while internal cytosines are RCY. Proflavines are PRF residues. The proton numberings are shown for proflavine as they do not follow the IUPAC numbering standards (as is shown in Figure 3.6)

Table A.1. Distance constraints applied to the AMBER model

Residue A	Residue A Name	Atom A Name	Residue B	Residue B Name	Atom B Name	Dist A (Å)	Dist B (Å)
1	RGX	H1'	1	RGX	H2'	1.8	3.3
1	RGX	H1'	1	RGX	H3'	1.8	5.0
1	RGX	H1'	1	RGX	H4'	1.8	5.0
1	RGX	H1'	1	RGX	H8	1.8	3.3
1	RGX	H1'	1	RGX	H5'	1.8	5.0
1	RGX	H1'	1	RGX	H5"	1.8	5.0
1	RGX	H2'	1	RGX	H3'	1.8	2.7
1	RGX	H2'	1	RGX	H4'	1.8	3.3
1	RGX	H2'	1	RGX	H8	1.8	5.0
1	RGX	H2'	1	RGX	H5'	1.8	5.0
1	RGX	H2'	1	RGX	H5"	1.8	5.0
1	RGX	H3'	1	RGX	H4'	1.8	2.7
1	RGX	H3'	1	RGX	H8	1.8	5.0

Table A.1 continued

1	RGX	H3'	1	RGX	H5'	1.8	3.3
1	RGX	H3'	1	RGX	H5"	1.8	3.3
1	RGX	H4'	1	RGX	H8	1.8	5.0
1	RGX	H4'	1	RGX	H5'	1.8	2.7
1	RGX	H4'	1	RGX	H5"	1.8	2.7
1	RGX	H5'	1	RGX	H5"	1.8	2.7
1	RGX	H1'	2	RCY	H1'	1.8	5.0
1	RGX	H1'	2	RCY	H4'	1.8	5.0
1	RGX	H1'	2	RCY	H5	1.8	5.0
1	RGX	H1'	2	RCY	H6	1.8	5.0
1	RGX	H1'	2	RCY	H5'	1.8	5.0
1	RGX	H1'	2	RCY	H5"	1.8	5.0
1	RGX	H2'	2	RCY	H4'	1.8	5.0
1	RGX	H2'	2	RCY	H5	1.8	5.0
1	RGX	H2'	2	RCY	H6	1.8	3.3
1	RGX	H2'	2	RCY	H5'	1.8	5.0
1	RGX	H2'	2	RCY	H5"	1.8	5.0
1	RGX	H3'	2	RCY	H5	1.8	5.0
1	RGX	H3'	2	RCY	H6	1.8	5.0
1	RGX	H8	2	RCY	H4'	1.8	5.0
1	RGX	H8	2	RCY	H5'	1.8	5.0
1	RGX	H8	2	RCY	H5"	1.8	5.0
2	RCY	H1'	2	RCY	H2'	1.8	5.0
2	RCY	H1'	2	RCY	H3'	1.8	5.0
2	RCY	H1'	2	RCY	H4'	1.8	5.0
2	RCY	H1'	2	RCY	H6	1.8	5.0
2	RCY	H1'	2	RCY	H5'	1.8	5.0
2	RCY	H1'	2	RCY	H5"	1.8	5.0
2	RCY	H2'	2	RCY	H4'	1.8	3.3
2	RCY	H2'	2	RCY	H6	1.8	2.7
2	RCY	H2'	2	RCY	H5'	1.8	5.0
2	RCY	H2'	2	RCY	H5"	1.8	5.0
2	RCY	H3'	2	RCY	H6	1.8	5.0
2	RCY	H4'	2	RCY	H6	1.8	5.0
2	RCY	H4'	2	RCY	H5'	1.8	2.7
2	RCY	H4'	2	RCY	H5"	1.8	2.7
2	RCY	H5	2	RCY	H6	1.8	2.7
2	RCY	H5	2	RCY	H5'	1.8	5.0
2	RCY	H5	2	RCY	H5"	1.8	5.0
2	RCY	H6	2	RCY	H5'	1.8	5.0
2	RCY	H6	2	RCY	H5"	1.8	5.0
2	RCY	H1'	3	RCY	H5	1.8	5.0
2	RCY	H1'	3	RCY	H6	1.8	5.0
2	RCY	H1'	3	RCY	H5'	1.8	5.0
2	RCY	H2'	3	RCY	H5	1.8	3.3

Table A.1 continued

2	RCY	H2'	3	RCY	H6	1.8	5.0
2	RCY	H3'	3	RCY	H5	1.8	5.0
2	RCY	H3'	3	RCY	H6	1.8	5.0
2	RCY	H5	3	RCY	H5	1.8	5.0
2	RCY	H5	3	RCY	H6	1.8	5.0
2	RCY	H6	3	RCY	H5	1.8	5.0
3	RCY	H1'	3	RCY	H2'	1.8	3.3
3	RCY	H1'	3	RCY	H3'	1.8	3.3
3	RCY	H1'	3	RCY	H4'	1.8	5.0
3	RCY	H1'	3	RCY	H5	1.8	5.0
3	RCY	H1'	3	RCY	H6	1.8	3.3
3	RCY	H2'	3	RCY	H3'	1.8	2.7
3	RCY	H2'	3	RCY	H4'	1.8	3.3
3	RCY	H2'	3	RCY	H5	1.8	3.3
3	RCY	H2'	3	RCY	H6	1.8	2.7
3	RCY	H2'	3	RCY	H5'	1.8	3.3
3	RCY	H3'	3	RCY	H4'	1.8	3.3
3	RCY	H3'	3	RCY	H5	1.8	5.0
3	RCY	H3'	3	RCY	H6	1.8	3.3
3	RCY	H3'	3	RCY	H5'	1.8	2.7
3	RCY	H4'	3	RCY	H6	1.8	5.0
3	RCY	H4'	3	RCY	H5'	1.8	2.7
3	RCY	H5	3	RCY	H6	1.8	2.7
3	RCY	H6	3	RCY	H5'	1.8	5.0
3	RCY	H1'	18	PRF	H6	1.8	5.0
3	RCY	H2'	18	PRF	H6	1.8	5.0
3	RCY	H3'	18	PRF	H6	1.8	5.0
3	RCY	H6	18	PRF	H5	1.8	5.0
4	RGY	H1'	4	RGY	H2'	1.8	5.0
4	RGY	H1'	4	RGY	H3'	1.8	3.3
4	RGY	H1'	4	RGY	H4'	1.8	5.0
4	RGY	H2'	4	RGY	H3'	1.8	2.7
4	RGY	H2'	4	RGY	H4'	1.8	5.0
4	RGY	H2'	4	RGY	H8	1.8	5.0
4	RGY	H3'	4	RGY	H4'	1.8	3.3
4	RGY	H3'	4	RGY	H8	1.8	3.3
4	RGY	H1'	5	RCY	H4'	1.8	5.0
4	RGY	H1'	5	RCY	H6	1.8	5.0
4	RGY	H1'	5	RCY	H5'	1.8	3.3
4	RGY	H2'	5	RCY	H2'	1.8	3.3
4	RGY	H2'	5	RCY	H6	1.8	5.0
4	RGY	H4'	5	RCY	H6	1.8	5.0
4	RGY	H1'	18	PRF	H6	1.8	5.0
4	RGY	H1'	18	PRF	H4	1.8	5.0
4	RGY	H2'	18	PRF	H6	1.8	5.0

Table A.1 continued

4	RGY	H2'	18	PRF	H7	1.8	5.0
4	RGY	H3'	18	PRF	H6	1.8	5.0
4	RGY	H4'	18	PRF	H6	1.8	5.0
4	RGY	H4'	18	PRF	H7	1.8	5.0
5	RCY	H1'	5	RCY	H2'	1.8	5.0
5	RCY	H1'	5	RCY	H3'	1.8	5.0
5	RCY	H1'	5	RCY	H4'	1.8	5.0
5	RCY	H1'	5	RCY	H5	1.8	5.0
5	RCY	H1'	5	RCY	H6	1.8	5.0
5	RCY	H1'	5	RCY	H5'	1.8	5.0
5	RCY	H2'	5	RCY	H6	1.8	2.7
5	RCY	H3'	5	RCY	H4'	1.8	3.3
5	RCY	H3'	5	RCY	H6	1.8	3.3
5	RCY	H3'	5	RCY	H5'	1.8	3.3
5	RCY	H4'	5	RCY	H6	1.8	5.0
5	RCY	H4'	5	RCY	H5'	1.8	2.7
5	RCY	H5	5	RCY	H6	1.8	2.7
5	RCY	H6	5	RCY	H5'	1.8	5.0
5	RCY	H1'	17	PRF	H6	1.8	5.0
5	RCY	H2'	17	PRF	H5	1.8	5.0
5	RCY	H2'	17	PRF	H6	1.8	5.0
5	RCY	H2'	17	PRF	H7	1.8	5.0
5	RCY	H3'	17	PRF	H6	1.8	5.0
5	RCY	H6	17	PRF	H5	1.8	5.0
6	RGY	H1'	6	RGY	H2'	1.8	2.7
6	RGY	H1'	6	RGY	H8	1.8	5.0
6	RGY	H2'	6	RGY	H3'	1.8	5.0
6	RGY	H2'	6	RGY	H4'	1.8	3.3
6	RGY	H2'	6	RGY	H8	1.8	3.3
6	RGY	H1'	7	RGY	H3'	1.8	3.3
6	RGY	H1'	7	RGY	H4'	1.8	2.7
6	RGY	H1'	7	RGY	H8	1.8	5.0
6	RGY	H2'	7	RGY	H8	1.8	5.0
6	RGY	H1'	17	PRF	H6	1.8	5.0
6	RGY	H1'	17	PRF	H4	1.8	5.0
6	RGY	H2'	17	PRF	H6	1.8	5.0
6	RGY	H2'	17	PRF	H7	1.8	5.0
6	RGY	H4'	17	PRF	H6	1.8	5.0
6	RGY	H4'	17	PRF	H7	1.8	5.0
6	RGY	H8	17	PRF	H6	1.8	5.0
6	RGY	H8	17	PRF	H4	1.8	5.0
7	RGY	H1'	7	RGY	H2'	1.8	2.7
7	RGY	H1'	7	RGY	H3'	1.8	3.3
7	RGY	H1'	7	RGY	H4'	1.8	5.0
7	RGY	H1'	7	RGY	H8	1.8	5.0

Table A.1 continued

7	RGY	H2'	7	RGY	H3'	1.8	2.7
7	RGY	H2'	7	RGY	H8	1.8	2.7
7	RGY	H3'	7	RGY	H4'	1.8	3.3
7	RGY	H3'	7	RGY	H8	1.8	3.3
7	RGY	H3'	7	RGY	H5'	1.8	2.7
7	RGY	H4'	7	RGY	H8	1.8	5.0
7	RGY	H4'	7	RGY	H5'	1.8	2.7
7	RGY	H8	7	RGY	H5'	1.8	5.0
7	RGY	H1'	8	RC3	H2'	1.8	5.0
7	RGY	H1'	8	RC3	H3'	1.8	5.0
7	RGY	H1'	8	RC3	H5	1.8	5.0
7	RGY	H3'	8	RC3	H6	1.8	5.0
7	RGY	H8	8	RC3	H2'	1.8	5.0
7	RGY	H8	8	RC3	H3'	1.8	5.0
7	RGY	H8	8	RC3	H4'	1.8	5.0
7	RGY	H8	8	RC3	H5	1.8	5.0
7	RGY	H8	8	RC3	H6	1.8	5.0
8	RC3	H1'	8	RC3	H2'	1.8	3.3
8	RC3	H1'	8	RC3	H3'	1.8	3.3
8	RC3	H1'	8	RC3	H4'	1.8	5.0
8	RC3	H1'	8	RC3	H5	1.8	5.0
8	RC3	H1'	8	RC3	H6	1.8	3.3
8	RC3	H1'	8	RC3	H5'	1.8	5.0
8	RC3	H1'	8	RC3	H5"	1.8	5.0
8	RC3	H2'	8	RC3	H3'	1.8	2.7
8	RC3	H2'	8	RC3	H4'	1.8	2.7
8	RC3	H2'	8	RC3	H5	1.8	5.0
8	RC3	H2'	8	RC3	H6	1.8	2.7
8	RC3	H3'	8	RC3	H4'	1.8	3.3
8	RC3	H3'	8	RC3	H5	1.8	5.0
8	RC3	H3'	8	RC3	H6	1.8	3.3
8	RC3	H4'	8	RC3	H4'	1.8	2.7
8	RC3	H4'	8	RC3	H5	1.8	5.0
8	RC3	H4'	8	RC3	H6	1.8	5.0
8	RC3	H4'	8	RC3	H5'	1.8	5.0
8	RC3	H4'	8	RC3	H5"	1.8	2.7
8	RC3	H5	8	RC3	H6	1.8	2.7
8	RC3	H5	8	RC3	H5'	1.8	5.0
8	RC3	H5'	8	RC3	H5"	1.8	2.7
8	RC3	H6	8	RC3	H5'	1.8	5.0
8	RC3	H6	8	RC3	H5"	1.8	5.0
9	RGX	H1'	9	RGX	H2'	1.8	3.3
9	RGX	H1'	9	RGX	H3'	1.8	5.0
9	RGX	H1'	9	RGX	H4'	1.8	5.0
9	RGX	H1'	9	RGX	H8	1.8	3.3

Table A.1 continued

9	RGX	H1'	9	RGX	H5'	1.8	5.0
9	RGX	H1'	9	RGX	H5"	1.8	5.0
9	RGX	H2'	9	RGX	H3'	1.8	2.7
9	RGX	H2'	9	RGX	H4'	1.8	3.3
9	RGX	H2'	9	RGX	H8	1.8	5.0
9	RGX	H2'	9	RGX	H5'	1.8	5.0
9	RGX	H2'	9	RGX	H5"	1.8	5.0
9	RGX	H3'	9	RGX	H4'	1.8	2.7
9	RGX	H3'	9	RGX	H8	1.8	5.0
9	RGX	H3'	9	RGX	H5'	1.8	3.3
9	RGX	H3'	9	RGX	H5"	1.8	3.3
9	RGX	H4'	9	RGX	H8	1.8	5.0
9	RGX	H4'	9	RGX	H5'	1.8	2.7
9	RGX	H4'	9	RGX	H5"	1.8	2.7
9	RGX	H5'	9	RGX	H5"	1.8	2.7
9	RGX	H1'	10	RCY	H1'	1.8	5.0
9	RGX	H1'	10	RCY	H4'	1.8	5.0
9	RGX	H1'	10	RCY	H5	1.8	5.0
9	RGX	H1'	10	RCY	H6	1.8	5.0
9	RGX	H1'	10	RCY	H5'	1.8	5.0
9	RGX	H1'	10	RCY	H5"	1.8	5.0
9	RGX	H2'	10	RCY	H4'	1.8	5.0
9	RGX	H2'	10	RCY	H5	1.8	5.0
9	RGX	H2'	10	RCY	H6	1.8	3.3
9	RGX	H2'	10	RCY	H5'	1.8	5.0
9	RGX	H2'	10	RCY	H5"	1.8	5.0
9	RGX	H3'	10	RCY	H5	1.8	5.0
9	RGX	H3'	10	RCY	H6	1.8	5.0
9	RGX	H8	10	RCY	H4'	1.8	5.0
9	RGX	H8	10	RCY	H5'	1.8	5.0
9	RGX	H8	10	RCY	H5"	1.8	5.0
10	RCY	H1'	10	RCY	H2'	1.8	5.0
10	RCY	H1'	10	RCY	H3'	1.8	5.0
10	RCY	H1'	10	RCY	H4'	1.8	5.0
10	RCY	H1'	10	RCY	H6	1.8	5.0
10	RCY	H1'	10	RCY	H5'	1.8	5.0
10	RCY	H1'	10	RCY	H5"	1.8	5.0
10	RCY	H2'	10	RCY	H4'	1.8	3.3
10	RCY	H2'	10	RCY	H6	1.8	2.7
10	RCY	H2'	10	RCY	H5'	1.8	5.0
10	RCY	H2'	10	RCY	H5"	1.8	5.0
10	RCY	H3'	10	RCY	H6	1.8	5.0
10	RCY	H4'	10	RCY	H6	1.8	5.0
10	RCY	H4'	10	RCY	H5'	1.8	2.7
10	RCY	H4'	10	RCY	H5"	1.8	2.7

Table A.1 continued

10	RCY	H5	10	RCY	H6	1.8	2.7
10	RCY	H5	10	RCY	H5'	1.8	5.0
10	RCY	H5	10	RCY	H5"	1.8	5.0
10	RCY	H6	10	RCY	H5'	1.8	5.0
10	RCY	H6	10	RCY	H5"	1.8	5.0
10	RCY	H1'	11	RCY	H5	1.8	5.0
10	RCY	H1'	11	RCY	H6	1.8	5.0
10	RCY	H1'	11	RCY	H5'	1.8	5.0
10	RCY	H2'	11	RCY	H5	1.8	3.3
10	RCY	H2'	11	RCY	H6	1.8	5.0
10	RCY	H3'	11	RCY	H5	1.8	5.0
10	RCY	H3'	11	RCY	H6	1.8	5.0
10	RCY	H5	11	RCY	H5	1.8	5.0
10	RCY	H5	11	RCY	H6	1.8	5.0
10	RCY	H6	11	RCY	H5	1.8	5.0
11	RCY	H1'	11	RCY	H2'	1.8	3.3
11	RCY	H1'	11	RCY	H3'	1.8	3.3
11	RCY	H1'	11	RCY	H4'	1.8	5.0
11	RCY	H1'	11	RCY	H5	1.8	5.0
11	RCY	H1'	11	RCY	H6	1.8	3.3
11	RCY	H2'	11	RCY	H3'	1.8	2.7
11	RCY	H2'	11	RCY	H4'	1.8	3.3
11	RCY	H2'	11	RCY	H5	1.8	3.3
11	RCY	H2'	11	RCY	H6	1.8	2.7
11	RCY	H2'	11	RCY	H5'	1.8	3.3
11	RCY	H3'	11	RCY	H4'	1.8	3.3
11	RCY	H3'	11	RCY	H5	1.8	5.0
11	RCY	H3'	11	RCY	H6	1.8	3.3
11	RCY	H3'	11	RCY	H5'	1.8	2.7
11	RCY	H4'	11	RCY	H6	1.8	5.0
11	RCY	H4'	11	RCY	H5'	1.8	2.7
11	RCY	H5	11	RCY	H6	1.8	2.7
11	RCY	H6	11	RCY	H5'	1.8	5.0
11	RCY	H1'	17	PRF	H3	1.8	5.0
11	RCY	H2'	17	PRF	H3	1.8	5.0
11	RCY	H3'	17	PRF	H3	1.8	5.0
11	RCY	H6	17	PRF	H1	1.8	5.0
12	RGY	H1'	12	RGY	H2'	1.8	5.0
12	RGY	H1'	12	RGY	H3'	1.8	3.3
12	RGY	H1'	12	RGY	H4'	1.8	5.0
12	RGY	H2'	12	RGY	H3'	1.8	2.7
12	RGY	H2'	12	RGY	H4'	1.8	5.0
12	RGY	H2'	12	RGY	H8	1.8	5.0
12	RGY	H3'	12	RGY	H4'	1.8	3.3
12	RGY	H3'	12	RGY	H8	1.8	3.3

Table A.1 continued

12	RGY	H1'	13	RCY	H4'	1.8	5.0
12	RGY	H1'	13	RCY	H6	1.8	5.0
12	RGY	H1'	13	RCY	H5'	1.8	3.3
12	RGY	H2'	13	RCY	H2'	1.8	3.3
12	RGY	H2'	13	RCY	H6	1.8	5.0
12	RGY	H4'	13	RCY	H6	1.8	5.0
12	RGY	H1'	17	PRF	H3	1.8	5.0
12	RGY	H1'	17	PRF	H4	1.8	5.0
12	RGY	H2'	17	PRF	H3	1.8	5.0
12	RGY	H2'	17	PRF	H2	1.8	5.0
12	RGY	H3'	17	PRF	H3	1.8	5.0
12	RGY	H4'	17	PRF	H3	1.8	5.0
12	RGY	H4'	17	PRF	H2	1.8	5.0
13	RCY	H1'	13	RCY	H2'	1.8	5.0
13	RCY	H1'	13	RCY	H3'	1.8	5.0
13	RCY	H1'	13	RCY	H4'	1.8	5.0
13	RCY	H1'	13	RCY	H5	1.8	5.0
13	RCY	H1'	13	RCY	H6	1.8	5.0
13	RCY	H1'	13	RCY	H5'	1.8	5.0
13	RCY	H2'	13	RCY	H6	1.8	2.7
13	RCY	H3'	13	RCY	H4'	1.8	3.3
13	RCY	H3'	13	RCY	H6	1.8	3.3
13	RCY	H3'	13	RCY	H5'	1.8	3.3
13	RCY	H4'	13	RCY	H6	1.8	5.0
13	RCY	H4'	13	RCY	H5'	1.8	2.7
13	RCY	H5	13	RCY	H6	1.8	2.7
13	RCY	H6	13	RCY	H5'	1.8	5.0
13	RCY	H1'	18	PRF	H3	1.8	5.0
13	RCY	H2'	18	PRF	H1	1.8	5.0
13	RCY	H2'	18	PRF	H3	1.8	5.0
13	RCY	H2'	18	PRF	H2	1.8	5.0
13	RCY	H3'	18	PRF	H3	1.8	5.0
13	RCY	H6	18	PRF	H1	1.8	5.0
14	RGY	H1'	14	RGY	H2'	1.8	2.7
14	RGY	H1'	14	RGY	H8	1.8	5.0
14	RGY	H2'	14	RGY	H3'	1.8	5.0
14	RGY	H2'	14	RGY	H4'	1.8	3.3
14	RGY	H2'	14	RGY	H8	1.8	3.3
14	RGY	H1'	15	RGY	H3'	1.8	3.3
14	RGY	H1'	15	RGY	H4'	1.8	2.7
14	RGY	H1'	15	RGY	H8	1.8	5.0
14	RGY	H2'	15	RGY	H8	1.8	5.0
14	RGY	H1'	18	PRF	H3	1.8	5.0
14	RGY	H1'	18	PRF	H4	1.8	5.0
14	RGY	H2'	18	PRF	H3	1.8	5.0

Table A.1 continued

14	RGY	H2'	18	PRF	H2	1.8	5.0
14	RGY	H4'	18	PRF	H3	1.8	5.0
14	RGY	H4'	18	PRF	H2	1.8	5.0
14	RGY	H8	18	PRF	H3	1.8	5.0
14	RGY	H8	18	PRF	H4	1.8	5.0
15	RGY	H1'	15	RGY	H2'	1.8	2.7
15	RGY	H1'	15	RGY	H3'	1.8	3.3
15	RGY	H1'	15	RGY	H4'	1.8	5.0
15	RGY	H1'	15	RGY	H8	1.8	5.0
15	RGY	H2'	15	RGY	H3'	1.8	2.7
15	RGY	H2'	15	RGY	H8	1.8	2.7
15	RGY	H3'	15	RGY	H4'	1.8	3.3
15	RGY	H3'	15	RGY	H8	1.8	3.3
15	RGY	H3'	15	RGY	H5'	1.8	2.7
15	RGY	H4'	15	RGY	H8	1.8	5.0
15	RGY	H4'	15	RGY	H5'	1.8	2.7
15	RGY	H8	15	RGY	H5'	1.8	5.0
15	RGY	H1'	16	RC3	H2'	1.8	5.0
15	RGY	H1'	16	RC3	H3'	1.8	5.0
15	RGY	H1'	16	RC3	H5	1.8	5.0
15	RGY	H3'	16	RC3	H6	1.8	5.0
15	RGY	H8	16	RC3	H2'	1.8	5.0
15	RGY	H8	16	RC3	H3'	1.8	5.0
15	RGY	H8	16	RC3	H4'	1.8	5.0
15	RGY	H8	16	RC3	H5	1.8	5.0
15	RGY	H8	16	RC3	H6	1.8	5.0
16	RC3	H1'	16	RC3	H2'	1.8	3.3
16	RC3	H1'	16	RC3	H3'	1.8	3.3
16	RC3	H1'	16	RC3	H4'	1.8	5.0
16	RC3	H1'	16	RC3	H5	1.8	5.0
16	RC3	H1'	16	RC3	H6	1.8	3.3
16	RC3	H1'	16	RC3	H5'	1.8	5.0
16	RC3	H1'	16	RC3	H5"	1.8	5.0
16	RC3	H2'	16	RC3	H3'	1.8	2.7
16	RC3	H2'	16	RC3	H4'	1.8	2.7
16	RC3	H2'	16	RC3	H5	1.8	5.0
16	RC3	H2'	16	RC3	H6	1.8	2.7
16	RC3	H3'	16	RC3	H4'	1.8	3.3
16	RC3	H3'	16	RC3	H5	1.8	5.0
16	RC3	H3'	16	RC3	H6	1.8	3.3
16	RC3	H4'	16	RC3	H4'	1.8	2.7
16	RC3	H4'	16	RC3	H5	1.8	5.0
16	RC3	H4'	16	RC3	H6	1.8	5.0
16	RC3	H4'	16	RC3	H5'	1.8	5.0
16	RC3	H4'	16	RC3	H5"	1.8	2.7

Table A.1 continued

16	RC3	H5	16	RC3	H6	1.8	2.7
16	RC3	H5	16	RC3	H5'	1.8	5.0
16	RC3	H5'	16	RC3	H5"	1.8	2.7
16	RC3	H6	16	RC3	H5'	1.8	5.0
16	RC3	H6	16	RC3	H5"	1.8	5.0

Table A2. Sugar pseudorotational constraints applied to the AMBER model

Residue	Residue Name	Constraint Type	Pseudo A (°)	Pseudo B (°)
1	RGX	PPA	126	198
2	RCY	PPA	0	54
3	RCY	PPA	126	198
4	RGY	PPA	0	54
5	RCY	PPA	126	198
6	RGY	PPA	0	54
7	RGY	PPA	126	198
8	RC3	PPA	126	198
9	RGX	PPA	126	198
10	RCY	PPA	0	54
11	RCY	PPA	126	198
12	RGY	PPA	0	54
13	RCY	PPA	126	198
14	RGY	PPA	0	54
15	RGY	PPA	126	198
16	RC3	PPA	126	198

Table A3. Backbone torsional constraints applied to the AMBER model

Residue	Residue Name	Torsion	Torsion A (°)	Torsion B (°)
1	RGX	ZETA	240	360
2	RCY	ZETA	240	360
3	RCY	ZETA	120	240
4	RGY	ZETA	240	360
5	RCY	ZETA	120	240
6	RGY	ZETA	240	360
7	RGY	ZETA	240	360
9	RGX	ZETA	240	360
10	RCY	ZETA	240	360
11	RCY	ZETA	120	240

Table A.3 continued

12	RGY	ZETA	240	360
13	RCY	ZETA	120	240
14	RGY	ZETA	240	360
15	RGY	ZETA	240	360
2	RCY	ALPHA	240	360
3	RCY	ALPHA	240	360
4	RGY	ALPHA	240	360
5	RCY	ALPHA	240	360
6	RGY	ALPHA	240	360
7	RGY	ALPHA	240	360
8	RC3	ALPHA	240	360
10	RCY	ALPHA	240	360
11	RCY	ALPHA	240	360
12	RGY	ALPHA	240	360
13	RCY	ALPHA	240	360
14	RGY	ALPHA	240	360
15	RGY	ALPHA	240	360
16	RC3	ALPHA	240	360

REFERENCES

1. Crick, F. (1970) "Central dogma of molecular biology." *Nature*, **227**: 561-563.

2. Voet, D., Somerset and Voet, J.G. (1995) *Biochemistry*. Second ed. John Wiley & Sons, Inc., Hoboken, NJ.

3. Miller, S.L. (1953) "A production of amino acids under possible primitive Earth conditions." *Science*, **117**: 528-529.

4. Oró, J. (1961) "Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions." *Nature*, **191**: 1193-1194.

5. Schopf, J.W. (ed.) (2002) *Life's origin: the beginnings of biological evolution*. University of California Press, Los Angeles, CA.

6. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) "Self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena." *Cell*, **31**: 147-157.

7. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) "The RNA moiety of ribonuclease-P is the catalytic subunit of the enzyme." *Cell*, **35**: 849-857.

8. Gilbert, W. (1986) "Origin of life - the RNA world." *Nature*, **319**: 618-618.

9. Breaker, R.R. and Joyce, G.F. (1994) "Inventing and improving ribozyme function-rational design versus iterative selection methods." *Trends Biotechnol.*, **12**: 268-275.

10. Müller, U.F. (2006) "Re-creating an RNA world." *Cell. Mol. Life Sci.*, **63**: 1278-1293.

11. Fedor, M.J. and Williamson, J.R. (2005) "The catalytic diversity of RNAs." *Nat. Rev. Mol. Cell Biol.*, **6**: 399-412.

12. Lincoln, T.A. and Joyce, G.F. (2009) Self-sustained replication of an RNA enzyme. *Science*, 10.1126/science.1167856.

13. Inoue, T. and Orgel, L.E. (1982) "Oligomerization of (guanosine 5'-phosphor)-2-methylimidazolide on poly(C) - an RNA-polymerase model." *J. Mol. Biol.*, **162**: 201-217.

14. Ferris, J.P., Hill, A.R., Liu, R.H. and Orgel, L.E. (1996) "Synthesis of long prebiotic oligonucleotides on mineral surfaces." *Nature*, **381**: 59-61.

15. Joshi, P.C., Pitsch, S. and Ferris, J.P. (2007) "Selectivity of montmorillonite catalyzed prebiotic reactions of D, L-nucleotides." *Origins Life Evol. B.*, **37**: 3-26.

16. Avery, O.T., MacLeod, C.M. and McCarty, M. (1944) "Studies on the chemical nature of the substance inducing transformation of pneumococcal types." *J. Exp. Med.*, **79**: 137-157.

17. Chargaff, E. (1950) "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation." *Experientia*, **6**: 201-209.

18. Hershey, A.D. and Chase, M. (1952) "Independent functions of viral protein and nucleic acid in growth of bacteriophage." *J Gen Physiol*, **36**: 39-56.

19. Watson, J.D. and Crick, F.H.C. (1953) "Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid." *Nature*, **171**: 737-738.

20. Waring, M.J. and Wakelin, L.P.G. (2003) *Forty years on, in small molecule DNA and RNA binders*. Wiley-VCH Verlag GmbH & Co, Berlin.

21. Bloomfield, V.A., Crothers, D.M., Tinoco, I. (2000) *Nucleic acids: structures, properties, and functions*. University Science Books, Sausalito, CA.

22. Franklin, R.E. and Gosling, R.G. (1953) "Molecular configuration in sodium thymonucleate." *Nature*, **171**: 740-741.

23. Wilkins, M.H.F., Stokes, A.R. and Wilson, H.R. (1953) "Molecular structure of deoxypentose nucleic acids." *Nature*, **171**: 738-740.

24. Arnott, S. and Hukins, D.W.L. (1973) "Refinement of structure of B-DNA and implications for analysis of X-ray-diffraction data from fibers of biopolymers." *J. Mol. Biol.*, **81**: 93-105.

25. Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., Vanboom, J.H., Vandermarel, G. and Rich, A. (1979) "Molecular-structure of a left-handed double helical DNA fragment at atomic resolution." *Nature*, **282**: 680-686.

26. Saenger, W. (1984) *Principles of nucleic acid structure*. Springer-Verlag, Berlin.

27. Hud, N.V. (2008) *Nucleic acid metal ion interactions*. First ed. Royal Society of Chemistry, Cambridge, UK.

28. Cerny, J., Kabelac, M. and Hobza, P. (2008) "Double-Helical -> Ladder Structural Transition in the B-DNA is Induced by a Loss of Dispersion Energy." *J. Am. Chem. Soc.*, **130**: 16055-16059.

29. Chiu, T.K., Kaczor-Grzeskowiak, M. and Dickerson, R.E. (1999) "Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A-T-T-C-G-C-G." *J. Mol. Biol.*, **292**: 589-608.

30. Horton, N.C. and Finzel, B.C. (1996) "The structure of an RNA/DNA hybrid: A substrate of the ribonuclease activity of HIV-1 reverse transcriptase." *J. Mol. Biol.*, **264**: 521-533.

31. Moore, M.H., Hunter, W.N., Destaintot, B.L. and Kennard, O. (1989) "DNA-drug interactions - the crystal-structure of d(CGATCG) complexed with daunomycin." *J. Mol. Biol.*, **206**: 693-705.

32. Reddy, B.S., Seshadri, T.P., Sakore, T.D. and Sobell, H.M. (1979) "Visualization of drug-nucleic acid interactions at atomic resolution. 5. structure of two aminoacridine-dinucleoside monophosphate crystalline complexes, proflavine-5-iodocytidylyl (3'-5') guanosine and acridine orange-5-iodocytidylyl (3'-5') guanosine." *J. Mol. Biol.*, **135**: 787-812.

33. Schelhorn, T., Kretz, S. and Zimmermann, H.W. (1992) "Reinvestigation of the binding of proflavine to DNA - is intercalation the dominant binding effect." *Cell. Mol. Biol.*, **38**: 345-365.

34. Wartell, R.M., Larson, J.E. and Wells, R.D. (1974) "Netropsin - specific probe for A-T regions of duplex deoxyribonucleic-acid." *J. Biol. Chem.*, **249**: 6719-6731.

35. Wade W.S., Mrksich, M. and Dervan, P.B. (1992) "Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif." *J. Am. Chem. Soc.*, **114**: 8783-8794.

36. Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. and Dervan, P.B. (1992) "Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin." *Proc. Natl. Acad. Sci. USA*, **89**: 7586-7590.

37. Hsu, C.F., Phillip, J.W., Trauger J.W., Farkas, M.E., Belitsky, J.M., Heckel, A., Olenyuk, B.Z., Puckett, J.W., Wang, C.C.C. and Dervan, P.B. (2007) "Completion of a programmable DNA-binding small molecule library." *Tetrahedron*, **63**: 6146-6151.

38. Chaires, J.B. (1997) "Energetics of drug-DNA interactions." *Biopolymers*, **44**: 201-215.

39. Harris, S.A. and Laughton, C.A. (2007) "A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy." *J Phys-Condens Mat*, **19**: 1-14.

40. Lerman, L.S. (1963) "Structure of DNA-acridine complex." *Proc. Natl. Acad. Sci. USA*, **49**: 94-102.

41. Lerman, L.S. (1961) "Structural considerations in interaction of DNA and acridines." *J. Mol. Biol.*, **3**: 18-30.

42. Neville, D.M. and Davies, D.R. (1966) "Interaction of acridine dyes with DNA: an X-Ray diffaraction and optical investigation." *J. Mol. Biol.*, **17**: 57-74.

43. Hollingshead, L.M. and Faulds, D. (1991) "Idarubicin - a review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in the chemotherapy of cancer." *Drugs*, **42**: 690-719.

44. Takusagawa, F., Carlson, R.G. and Weaver, R.F. (2001) "Anti-leukemia selectivity in actinomycin analogues." *Bioorg. Med. Chem.*, **9**: 719-725.

45. Crothers, D.M. (1968) "Calculation of binding isotherms for heterogeneous polymers." *Biopolymers*, **6**: 575-584.

46. Voet, D. (1977) "Intercalation complexes of DNA." *Nature*, **269**: 285-286.

47. Berman, H.M., Stallings, W., Carrell, H.L., Glusker, J.P., Neidle, S., Taylor, G. and Achari, A. (1979) "Molecular and crystal-structure of an intercalation complex - proflavine-cytidylyl-(3',5')-guanosine." *Biopolymers*, **18**: 2405-2429.

48. Shieh, H.S., Berman, H.M., Dabrow, M. and Neidle, S. (1980) "Structure of drug-deoxydinucleoside phosphate complex - generalized conformational behavior of intercalation complexes with RNA and DNA fragments." *Nucleic Acids Res.*, **8**: 85-97.

49. Dearing, A., Weiner, P. and Kollman, P.A. (1981) "Molecular mechanical studies of proflavine and acridine-orange intercalation." *Nucleic Acids Res.*, **9**: 1483-1497.

50. Neidle, S. and Abraham, Z. (1984) "Structural and sequence-dependent aspects of drug intercalation into nucleic-acids." *CRC Crit. Rev. Biochem.*, **17**: 73-121.

51. Friedman, R.A.G. and Manning, G.S. (1984) "Poly-electrolyte effects on site-binding equilibria with application to the intercalation of drugs into DNA." *Biopolymers*, **23**: 2671-2714.

52. Rao, S.N. and Kollman, P.A. (1987) "Molecular mechanical simulations on double intercalation of 9-amino acridine into d(CGCGCGC)•d(GCGCGCG) - analysis of the physical basis for the neighbor-exclusion principle." *Proc. Natl. Acad. Sci. USA*, **84**: 5735-5739.

53. Hud, N.V. and Anet, F.A.L. (2000) "Intercalation-mediated synthesis and replication: a new approach to the origin of life." *J. Theor. Biol.*, **205**: 543-562.

54. Hud, N.V., Jain, S.S., Li, X.H. and Lynn, D.G. (2007) "Addressing the problems of base pairing and strand cyclization in template-directed synthesis - A case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA." *Chem. Biodiversity*, **4**: 768-783.

55. Sawada, T., Yoshizawa, M., Sato, S. and Fujita, M. (2009) Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts. *Nat. Chem.*, 10.1038/NCHEM.100.

56. Hoogsteen, K. (1963) "Crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine." *Acta Crystallographica*, **16**: 907-916.

57. Ren, J.S. and Chaires, J.B. (1999) "Sequence and structural selectivity of nucleic acid binding ligands." *Biochemistry*, **38**: 16067-16075.

58. Polak, M. and Hud, N.V. (2002) "Complete disproportionation of duplex poly(dT)•poly(dA) into triplex poly(dT)•poly(dA)-poly(dT) and poly(dA) by coralyne." *Nucleic Acids Res.*, **30**: 983-992.

59. Jain, S.S., Polak, M. and Hud, N.V. (2003) "Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne-driven duplex disproportionation." *Nucleic Acids Res.*, **31**: 4608-4615.

60. Persil, O., Santai, C.T., Jain, S.S. and Hud, N.V. (2004) "Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding." *J. Am. Chem. Soc.*, **126**: 8644-8645.

61. Jain, S.S., Anet, F.A.L., Stahle, C.J. and Hud, N.V. (2004) "Enzymatic behavior by intercalating molecules in a template-directed ligation reaction." *Angew Chem Int Edit*, **43**: 2004-2008.

62. Lohrmann, R., Bridson, P.K., Bridson, P.K. and Orgel, L.E. (1980) "Efficient metal-ion catalyzed template-directed oligonucleotide synthesis." *Science*, **208**: 1464-1465.

63. Kanaya, E. and Yanagawa, H. (1986) "Template-directed polymerization of oligoadenylates using cyanogen-bromide." *Biochemistry*, **25**: 7423-7430.

64. Ertem, G. and Ferris, J.P. (1997) "Template-directed synthesis using the heterogeneous templates produced by montmorillonite catalysis. A possible bridge between the prebiotic and RNA worlds." *J. Am. Chem. Soc.*, **119**: 7197-7201.

65. Huang, W.H. and Ferris, J.P. (2003) "Synthesis of 35-40 mers of RNA oligonucleotides from unblocked monomers. A simple approach to the RNA world." *Chem Commun*: 1458-1459.

66. Bean, H.D., Anet, F.A.L., Gould, I.R. and Hud, N.V. (2006) "Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA." *Origins Life Evol. B.*, **36**: 39-63.

67. Kolb, V.M., Dworkin, J.P. and Miller, S.L. (1994) "Alternative bases in the RNA world - the prebiotic synthesis of urazole and its ribsides." *J. Mol. Evol.*, **38**: 549-557.

68. Eschenmoser, A. (1999) "Chemical etiology of nucleic acid structure." *Science*, **284**: 2118-2124.

69. Nelson, K.E., Levy, M. and Miller, S.L. (2000) "Peptide nucleic acids rather than RNA may been the first genetic molecule." *Proc. Natl. Acad. Sci. USA*, **97**: 3868-3871.

70. Waring, M.J. (1965) "Complex formation between ethidium bromide and nucleic acids." *J. Mol. Biol.*, **13**: 269-282.

71. Baguley, B.C. (1991) "DNA intercalating antitumor agents." *Anti-Cancer Drug Des*, **6**: 1-35.

72. Ihmels, H. and Otto, D. (2005) "Intercalation of organic dye molecules into double-stranded DNA general principles and recent developments." *Top. Curr. Chem.*, **258**: 161-204.

73. Horowitz, E.D. and Hud, N.V. (2006) "Ethidium and proflavine binding to a 2',5'-linked RNA duplex." *J. Am. Chem. Soc.*, **128**: 15380-15381.

74. Jain, S.C. and Sobell, H.M. (1984) "Visualization of drug-nucleic acid interactions at atomic resolution. 8. structures of two ethidium dinucleoside monophosphate crystalline complexes containing ethidium - cytidylyl (3'-5') guanosine." *J. Biomol. Struct. Dyn.*, **1**: 1179-1194.

75. Cohn, W.E. (1957) "Methods of isolation and characterization of mononucleotides and polynucleotides by ion exchange chromatography." *Methods Enzymol.*, **3**: 724-743.

76. Cavaluzzi, M.J. and Borer, P.N. (2004) "Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA." *Nucleic Acids Res.*, **32**: e13.

77. Qu, X.G. and Chaires, J.B. (2000) "Analysis of drug-DNA binding data." *Method Enzymol*, **321**: 353-369.

78. Tanaka, S., Baba, Y., Kagemoto, A. and Fujishiro, R. (1981) "Thermodynamic studies on poly(A)•poly(U) duplex-dye systems." *Makromol. Chem.*, **182**: 2837-2843.

79. Wilson, W.D., Krishnamoorthy, C.R., Wang Y.H. and Smith, J.C. (1985) "Mechanism of intercalation-ion effects on the equilibrium and kinetic constants for the interaction of propidium and ethidium with DNA." *Biopolymers*, **24**: 1941-1961.

80. Manning, G.S. (1978) "Molecular theory of polyelectrolyte solutions with applications to electrostatic properties of polynucleotides." *Q. Rev. Biophys.*, **11**: 179-246.

81. Record, M.T., Anderson, C.F. and Lohman, T.M. (1978) "Thermodynamic analysis of ion effects on binding and conformational equilibria of proteins and nucleic acids-roles of ion association or release, screening, and ion effects on water activity." *Q. Rev. Biophys.*, **11**: 103-178.

82. Peacocke, A.R. and Skerrett, J.N.H. (1956) "The interaction of aminoacridines with nucleic acids." *Trans. Faraday Soc*, **52**: 261-279.

83. Jang, Y.J., Kwon, B.-H., Bae, C.H., Seo, M.S., Nam, W. and Kim, S.K. (2008) "Intercalation of bulky D,D- and L,L-bis-Ru(II) complex between DNA base pairs." *J. Inorg. Biochem.*, **102**: 1885-1891.

84. Ardhammar, M.N., B., Kurucsev, T. (2000) *Circular dichroism: principles and applications*. Second ed. John Wiley & Sons, Inc., Hoboken, NJ.

85. Job, P. (1928) "Studies on the formation of complex minerals in solution and on their stability." *Ann Chim France*, **9**: 113-203.

86. Sawai, H., Kuroda, K., Seki, J. and Ozaki, H. (1996) "Conformational and stacking properties of 3'-5' and 2'-5' linked oligoribonucleotides studied by CD." *Biopolymers*, **39**: 173-182.

87. Premraj, B.J., Patel, P.K., Kandimalla, E.R., Agrawal, S., Hosur, R.V. and Yathindra, N. (2001) "NMR structure of a 2',5' RNA favors A type duplex with

compact C2' endo nucleotide repeat." *Biochem. Biophys. Res. Commun.*, **283**: 537-543.

88. Qu, X.G. and Chaires, J.B. (2001) "Hydration changes for DNA intercalation reactions." *J. Am. Chem. Soc.*, **123**: 1-7.

89. Portugal, J., Cashman, D.J., Trent, J.O., Ferrer-Miralles, N., Przewloka, T., Fokt, I., Priebe, W. and Chaires, J.B. (2005) "A new bisintercalating anthracycline with picomolar DNA binding affinity." *J. Med. Chem.*, **48**: 8209-8219.

90. Marin, V., Hansen, H.F., Koch, T. and Armitage, B.A. (2004) "Effect of LNA Modifications on Small Molecule Binding to Nucleic Acids." *J. Biomol. Struct. Dyn.*, **21**: 841-850.

91. Leumann, C.J. (2002) "DNA analogues: From supramolecular principles to biological properties." *Bioorg. Med. Chem.*, **10**: 841-854.

92. Wilson, W.D., Tanious, F.A., Mizan, S., Yao, S.J., Kiselyov, A.S., Zon, G. and Strekowski, L. (1993) "DNA triple-helix specific intercalators as antigene enhancers - unfused aromatic cations." *Biochemistry*, **32**: 10614-10621.

93. Zunino, F., Dimarco, A., Zaccara, A. and Luoni, G. (1974) "Inhibition of RNA-polymerase by daunomycin." *Chem.-Biol. Interact.*, **9**: 25-36.

94. Poot, M., Hiller, K.H., Heimpel, S. and Hoehn, H. (1995) "Distinct patterns of cell-cycle disturbance elicited by compounds interfering with DNA topoisomerase-I and topoisomerase-II activity." *Exp. Cell Res.*, **218**: 326-330.

95. Popanda, O. and Thielmann, H.W. (1992) "The function of DNA topoisomerases in UV-induced DNA excision repair - studies with specific inhibitors in permeabilized human fibroblasts." *Carcinogenesis*, **13**: 2321-2328.

96. Williams, L.D., Egli M., Gao Q., Rich A. (1992) *Structure and function*. Adenine Press, Schenectady, NY.

97. Aggarwal, A., Islam, S.A., Kuroda, R. and Neidle, S. (1984) "X-Ray crystallographic analysis of a ternary intercalation complex between proflavine and the dinucleoside monophosphates CpA and UpG." *Biopolymers*, **23**: 1025-1041.

98. Usher, D.A. (1972) "RNA double helix and evolution of 3',5' linkage." *Nature-New Biol*, **235**: 207-208.

99. Kierzek, R., He, L.Y. and Turner, D.H. (1992) "Association of 2'-5' oligoribonucleotides." *Nucleic Acids Res.*, **20**: 1685-1690.

100. Giannaris, P.A. and Damha, M.J. (1993) "Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssDNA." *Nucleic Acids Res.*, **21**: 4742-4749.

101. Hannoush, R.N. and Damha, M.J. (2001) "Remarkable stability of hairpins containing 2',5'-linked RNA loops." *J. Am. Chem. Soc.*, **123**: 12368-12374.

102. Plevnik, M., Gdaniec, Z. and Plavec, J. (2005) "Solution structure of a modified 2',5'-linked RNA hairpin involved in an equilibrium with duplex." *Nucleic Acids Res.*, **33**: 1749-1759.

103. Horowitz, E.D., Lilavivat, S., Holladay, B.W., Germann, M.W. and Hud, N.V. (2009) Solution structure and thermodynamics of 2′,5′ RNA intercalation. *J. Am. Chem. Soc.*, 10.1021/ja810068e.

104. Piotto, M., Saudek, V. and Sklenar, V. (1992) "Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions." *J. Biomol. NMR*, **2**: 661-665.

105. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Debolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995) "AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules." *Comput. Phys. Commun.*, **91**: 1-41.

106. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) "Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers." *Biophys. J.*, **92**: 3817-3829.

107. Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) "Development and testing of a general AMBER force field." *J. Comput. Chem.*, **25**: 1157-1174.

108. Becke, A.D. (1993) "A new mixing of Hartree-Fock and local density-functional theories." *J. Chem. Phys.*, **98**: 1372-1377.

109. Mazzini, S., Mondelli, R. and Ragg, E. (1998) "Structure and dynamics of intercalation complexes of anthracyclines with d(CGATCG)(2) and d(CGTACG)(2). 2D-H-1 and P-31 NMR investigations." *J Chem Soc Perk T 2*: 1983-1991.

110. Tsui, V. and Case, D.A. (2000) "Molecular dynamics simulations of nucleic acids with a generalized born solvation model." *J. Am. Chem. Soc.*, **122**: 2489-2498.

111. Humphrey, W., Dalke, A. and Schulten, K. (1996) "VMD: Visual molecular dynamics." *J. Mol. Graphics*, **14**: 33-38.

112. Haq, I., Jenkins, T.C., Chowdhry, B.Z., Ren, J.S. and Chaires, J.B. (2000) "Parsing free energies of drug-DNA interactions." *Methods Enzymol.*, **323**: 373-405.

113. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. John Wiley & Sons, Inc., Hoboken, NJ.

114. Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1984) "Interactions of antitumor drugs with natural DNA - H1-NMR study of binding mode and kinetics." *J. Med. Chem.*, **27**: 450-465.

115. Lee, J., Guelev, V., Sorey, S., Hoffman, D.W. and Iverson, B.L. (2004) "NMR structural analysis of a modular threading tetra intercalator bound to DNA." *J. Am. Chem. Soc.*, **126**: 14036-14042.

116. Reinhardt, C.G. and Krugh, T.R. (1978) "Comparative-study of ethidium-bromide complexes with dinucleotides and DNA - direct evidence for intercalation and nucleic-acid sequence preferences." *Biochemistry*, **17**: 4845-4854.

117. Robinson, H., Jung, K.E., Switzer, C. and Wang, A.H.J. (1995) "DNA with 2'-5' phosphodiester bonds forms a duplex structure in the A-type conformation." *J. Am. Chem. Soc.*, **117**: 837-838.

118. Premraj, B.J., Raja, S., Bhavesh, N.S., Shi, K., Hosur, R.V., Sundaralingam, M. and Yathindra, N. (2004) "Solution structure of 2',5' d(G4C4) - Relevance to topological restrictions and nature's choice of phosphodiester links." *Eur. J. Biochem.*, **271**: 2956-2966.

119. Krishnan, R. and Seshadri, T.P. (1994) "Crystal-structure of guanylyl-2',5'-cytidine dihydrate-an analog of msDNA-RNA junction in Stigmatella-aurantiaca." *Biopolymers*, **34**: 1637-1646.

120. Gilbert, D.E. and Feigon, J. (1991) "The DNA-sequence at echinomycin binding-sites determines the structural-changes induced by drug-binding - NMR-studies of echinomycin binding to [d(ACGTACGT)]2 and [d(TCGATCGA)]2." *Biochemistry*, **30**: 2483-2494.

121. Spielmann, H.P., Wemmer, D.E. and Jacobsen, J.P. (1995) "Solution structure of a DNA complex with the fluorescent bis-intercalator TOTO determined by NMR-spectroscopy." *Biochemistry*, **34**: 8542-8553.

122. Adams, A. (2002) "Crystal structures of acridines complexed with nucleic acids." *Curr. Med. Chem.*, **9**: 1667-1675.

123. Gorenstein, D.G. (1992) "P-31 NMR of DNA." *Methods Enzymol.*, **211**: 254-286.

124. Williams, H.E.L. and Searle, M.S. (1999) "Structure, dynamics and hydration of the nogalamycin-d(ATGCAT)(2) complex determined by NMR and molecular dynamics simulations in solution." *J. Mol. Biol.*, **290**: 699-716.

125. Gorenstein, D.G., Goldfield, E.M. (1984), *Phosphorous-31 NMR: principles and applications*. Academic Press, Burlington, MA, pp. 299-316.

126. Evans, J.N.S. (1995) *Biomolecular NMR spectroscopy*. Oxford University Press, New York, NY.

127. Patel, D.J. and Shen, C. (1978) "Sugar pucker geometries at intercalation site of propidium diiodide into miniature RNA and DNA duplexes in solution." *Proc. Natl. Acad. Sci. USA*, **75**: 2553-2557.

128. Chaires, J.B. (2006) "A thermodynamic signature for drug-DNA binding mode." *Arch. Biochem. Biophys.*, **453**: 26-31.

129. Krugh, T.R., Wittlin, F.N. and Cramer, S.P. (1975) "Ethidium Bromide-Dinucleotide Complexes. Evidence for Intercalation and Sequence Preference in Binding to Double-Stranded Nucleic Acids." *Biopolymers*, **14**: 197-210.

130. Broyde, S. and Hingerty, B. (1979) "Conformational origin of the pyrimidine (3'-5') purine base sequence preference for intercalation into RNAs." *Biopolymers*, **18**: 2905-2910.

131. Premraj, B.J., Raja, S. and Yathindra, N. (2002) "Structural basis for the unusual properties of 2',5' nucleic acids and their complexes with RNA and DNA." *Biophys. Chem.*, **95**: 253-272.

132. Jain, S.C., Tsai, C. and Sobell, H.M. (1977) "Visualization of drug-nucleic acid interactions at atomic resolution. 2. structure of an ethidium-dinucleoside monophosphate crystalline complex, ethidium - 5-iodocytidylyl (3'-5') guanosine." *J. Mol. Biol.*, **114**: 317-331.

133. Tsai, C., Jain, S.C. and Sobell, H.M. (1977) "Visualization of drug-nucleic acid interactions at atomic resolution. 1. structure of an ethidium-dinucleoside monophosphate crystalline complex, ethidium - 5-iodouridylyl (3'-5') adenosine." *J. Mol. Biol.*, **114**: 301-315.

134. Prabhakaran, M. and Harvey, S.C. (1988) "Molecular-dynamics of structural transitions and intercalation in DNA." *Biopolymers*, **27**: 1239-1248.

135. Trieb, M., Rauch, C., Wibowo, F.R., Wellenzohn, B. and Liedl, K.R. (2004) "Cooperative effects on the formation of intercalation sites." *Nucleic Acids Res.*, **32**: 4696-4703.

136. Polak, M., Manoharan, M., Inamati, G.B. and Plavec, J. (2003) "Tuning of conformational preorganization in model 2',5'- and 3',5'-linked oligonucleotides by 3'- and 2'-O-methoxyethyl modification." *Nucleic Acids Res.*, **31**: 2066-2076.

137. Tian, Y., Kayatta, M., Shultis, K., Gonzalez, A., Mueller, L.J., Hatcher, M.E. (2008) 31P NMR Investigation of backbone dynamics in DNA binding sites. *J. Phys. Chem. B*, 10.1021/jp711203m.

138. Rehm, T. and Schmuck, C. (2008) "How to achieve self-assembly in polar solvents based on specific interactions? some general guidelines." *Chem Commun*: 801-813.

139. Costantino, L., Guarino, G., Ortona, O. and Vitagliano, V. (1984) "Acridine-orange association equilibrium in aqueous-solution." *J. Chem. Eng. Data*, **29**: 62-66.

140. Ortona, O., Costantino, L., Dellavolpe, C. and Vitagliano, V. (1990) "Stacking equilibria of proflavine in various solutions." *J. Mol. Liq.*, **45**: 201-211.

141. Persil, Ö. and Hud, N.V. (2007) "Harnessing DNA intercalation." *Trends Biotechnol.*, **25**: 433-436.

142. Çetinkol, Ö.P. and Hud, N.V. (2009) "Molecular recognition of poly(A) by small ligands: an alternative method of analysis reveals nanomolar, cooperative and shape-selective binding." *Nucleic Acids Res.*, **37**: 611-621.

143. Joyce, G.F. (1989) "RNA evolution and the origins of life." *Nature*, **338**: 217-224.

144. Gesteland, R.F., Cech, T.R., Atkins, J.F. (2006) *The RNA world*. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring harbor, NY.

145. Ellington, A.D., Chen, X., Robertson, M. and Syrett, A. (2009) "Evolutionary origins and directed evolution of RNA." *Int. J. Biochem. Cell Biol.*, **41**: 254-265.

146. Heuberger, B.D. and Switzer, C. (2008) "A pre-RNA candidate revisited: both enantiomers of flexible nucleoside triphosphates are DNA polymerase substrates." *J. Am. Chem. Soc.*, **130**: 412-413.

147. Mittapalli, G.K., Reddy, K.R., Xiong, H., Munoz, O., Han, B., De Riccardis, F., Krishnamurthy, R. and Eschenmoser, A. (2007) "Mapping the landscape of potentially primordial informational oligonucleotides: oligodipeptides and

oligodipeptoids tagged with triazines as recognition elements." *Angew. Chem., Int. Ed. Engl.*, **46**: 2470-2477.

148. Benner, S.A., Ricardo, A. and Carrigan, M.A. (2004) "Is there a common chemical model for life in the universe?" *Curr. Opin. Chem. Biol.*, **8**: 672-689.

149. Benner, S.A. (2004) "Understanding nucleic acids using synthetic chemistry." *Acc. Chem. Res.*, **37**: 784-797.

150. Leitzel, J.C. and Lynn, D.G. (2001) "Template-directed ligation: From DNA towards different versatile templates." *Chem. Rec.*, **1**: 53-62.

151. Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F. and Di Mauro, E. (2007) "Nucleoside phosphorylation by phosphate minerals." *J. Biol. Chem.*, **282**: 16729-16735.

152. Anastasi, C., Buchet, F.F., Crowe, M.A., Parkes, A.L., Powner, M.W., Smith, J.M. and Sutherland, J.D. (2007) "RNA: prebiotic product, or biotic invention?" *Chem. Biodiversity*, **4**: 721-739.

153. Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F. and Deamer, D. (2008) "Lipid-assisted synthesis of RNA-like polymers from mononucleotides." *Origins Life Evol. B.*, **38**: 57-74.

154. Huang, W. and Ferris, J.P. (2006) "One-step, regioselective synthesis of up to 50-mers of RNA oligonucleotides by montmorillonite catalysis." *J. Am. Chem. Soc.*, **128**: 8914-8919.

155. Joyce, G.F., Inoue, T. and Orgel, L.E. (1984) "Non-enzymatic template-directed synthesis on RNA random copolymers - poly(C,U) templates." *J. Mol. Biol.*, **176**: 279-306.

156. Weimann, B.J., Lohrmann, R., Orgel, L.E., Schneide.H and Sulston, J.E. (1968) "Template-directed synthesis with adenosine-5'-phosphorimidazolide." *Science*, **161**: 387.

157. Smietana, M. and Kool, E.T. (2002) "Efficient and simple solid-phase synthesis of short cyclic oligodeoxynucleotides bearing a phosphorothioate linkage." *Angew Chem Int Edit*, **41**: 3704-3707.

158. Kawamura, K. and Okamoto, F. (2001) "Cyclization and dimerization of hexanucleotides containing guanine and cytosine with water-soluble carbodiimide." *Viva Origino*, **29**: 162-167.

159. Zielinski, W.S. and Orgel, L.E. (1987) "Oligoaminonucleoside phosphoramidites. Oligomerization of dimers of 3'-amino-3'-deoxy-nucleotides (GC and CG) in aqueous-solution." *Nucleic Acids Res.*, **15**: 1699-1715.

160. Jacobson, H. and Stockmayer, W.H. (1950) "Intramolecular reaction in polycondensations. 1. the theory of linear systems." *J. Chem. Phys.*, **18**: 1600-1606.

161. Mills, J.B., Vacano, E. and Hagerman, P.J. (1999) "Flexibility of single-stranded DNA: Use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA)." *J. Mol. Biol.*, **285**: 245-257.

162. Berge, T., Jenkins, N.S., Hopkirk, R.B., Waring, M.J., Edwardson, J.M. and Henderson, R.M. (2002) "Structural perturbations in DNA caused by bis-intercalation of ditercalinium visualised by atomic force microscopy." *Nucleic Acids Res.*, **30**: 2980-2986.

163. Huang, K.S., Haddadin, M.J., Olmstead, M.M. and Kurth, M.J. (2001) "Synthesis and reactions of some heterocyclic azacyanines." *J. Org. Chem.*, **66**: 1310-1315.

164. Lepecq, J.B. and Paoletti, C. (1967) "A fluorescent complex between ethidium bromide and nucleic acids - physical-chemical characterization." *J. Mol. Biol.*, **27**: 87-106.

165. Çetinkol, Ö.P., Engelhart, A.E., Nanjunda, R.K., Wilson, W.D. and Hud, N.V. (2008) "Submicromolar, selective G-quadruplex ligands from one pot: thermodynamic and structural studies of human telomeric DNA binding by azacyanines." *ChemBioChem*, **9**: 1889-1892.

166. James, K.D. and Ellington, A.D. (1997) "Surprising fidelity of template-directed chemical ligation of oligonucleotides." *Chem. Biol.*, **4**: 595-605.

167. Sievers, D. and von Kiedrowski, G. (1994) "Self replication of complementary nucleotide-based oligonucleotides." *Nature*, **369**: 221-224.

168. von Kiedrowski, G. (1986) "A self-replicating hexadeoxynucleotide." *Angew. Chem., Int. Ed. Engl.*, **25**: 932-935.

169. Orgel, L.E. and Lohrmann, R. (1974) "Prebiotic chemistry and nucleic acid replication." *Acc. Chem. Res.*, **7**: 368-377.

170. Eschenmoser, A. and Loewenthal, E. (1992) "Chemistry of potentially prebiological natural products." *Chem. Soc. Rev.*, **21**: 1-16.

171. Schoning, K.U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. and Eschenmoser, A. (2000) "Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3'-2') oligonucleotide system." *Science*, **290**: 1347-1351.

172. Eschenmoser, A. (2005) "Searching for nucleic acid alternatives." *Chimia*, **59**: 836-850.

173. Groebke, K., Hunziker, J., Fraser, W., Peng, L., Diederichsen, U., Zimmermann, K., Holzner, A., Leumann, C. and Eschenmoser, A. (1998) "Why pentose- and not hexose-nucleic acids? Purine-purine pairing in homo-DNA: guanine, isoguanine, 2,6-diaminopurine, and xanthine." *Helv. Chim. Acta*, **81**: 375-474.

174. Battersby, T.R., Albalos, M. and Friesenhahn, M.J. (2007) "An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids." *Chem. Biol.*, **14**: 525-531.

175. Heuberger, B.D. and Switzer, C. (2008) "An alternative nucleobase code: characterization of purine-purine DNA double helicies bearing guanine-isoguanine and diaminopurine 7-deaza-xanthine base pairs." *ChemBioChem*, **9**: 2779-2783.

176. Rich, A. (1958) "Formation of two- and three-stranded helical molecules by polyinosinic acid and polyadenylic acid." *Nature*, **454**: 122-125.

177. Mansy, S.S., Schrum, J.P., Krishnamurth, M., Tobe, S., Treco, D.A. and Szostak, J.W. (2008) "Template-directed synthesis of a genetic polymer in a model protocell." *Nature*, **454**: 122-126.

178. Nakata, M., Zanchetta, G., Chapman, B.D., Jones, C.D., Cross, J.O., Pindak, R., Bellini, T. and Clark, N.A. (2007) "End-to-end stacking and liquid crystal condensation of 6-to 20-base pair DNA duplexes." *Science*, **318**: 1276-1279.

179. Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F. and Di Mauro, E. (2007) "Formamide as the main building block in the origin of nucleic acids." *BMC Evol. Biol.*, **7**: S1.

180. Saladino, R., Crestini, C., Ciciriello, F., Costanzo, G. and Di Mauro, E. (2007) "Formamide chemistry and the origin of informational polymers." *Chem. Biodiversity*, **4**: 694-720.

181. Mizuno, T. and Weiss, A.H. (1974) "Synthesis and utilization of formose sugars." *Adv. Carbohydr. Chem. Biochem.*, **29**: 173-227.

182. Decker, P., Schweer, H. and Pohlmann, R. (1982) "Bioids.10. identification of formose sugars, presumable prebiotic metabolites, using capillary gas-

chromatography gas-chromatography mass-spectrometry of normal-butoxime trifluoroacetates on Ov-225." *J. Chromatogr.*, **244**: 281-291.

183. Dworkin, J.P., Lazcano, A. and Miller, S.L. (2003) "The roads to and from the RNA world." *J. Theor. Biol.*, **222**: 127-134.

184. Jungmann, O., Wippo, H., Stanek, M., Huynh, H.K., Krishnamurthy, R. and Eschenmoser, A. (1999) "Promiscuous Watson-Crick cross-pairing within the family of pentopyranosyl (4'-2') oligonucleotides." *Org. Lett.*, **1**: 1527-1530.

185. Schlegel, M.K., Peritz, A.E., Kittigowittana, K., Zhang, L.L. and Meggers, E. (2007) "Duplex formation of the simplified nucleic acid GNA." *ChemBioChem*, **8**: 927-932.

186. Joyce, G.F., Schwartz, A.W., Miller, S.L. and Orgel, L.E. (1987) "The case for an ancestarl genetic system involving simple analogs of the nucleotides." *Proc. Natl. Acad. Sci. USA*, **84**: 4398-4402.

187. Schneider, K.C. and Benner, S. (1990) "Oligonucleotides containing flexible nucleoside analogues." *J. Am. Chem. Soc.*, **112**: 453-455.

188. Wasner, M., Arion, D., Borkow, G., Noronha, A., Uddin, A.H., Parniak, M.A. and Damha, M.J. (1998) "Physicochemical and biochemical properties of 2',5'-linked RNA and 2',5'-RNA : 3',5'-RNA "hybrid" duplexes." *Biochemistry*, **37**: 7478-7486.

189. Nielsen, K.E. and Spielmann, H.P. (2005) "The structure of a mixed LNA/DNA : RNA duplex is driven by conformational coupling between LNA and deoxyribose residues as determined from C-13 relaxation measurements." *J. Am. Chem. Soc.*, **127**: 15273-15282.

190. Jayakumar, H.K., Buckingham, J.L., Brazier, J.A., Berry, N.G., Cosstick, R. and Fisher, J. (2007) "NMR studies of the conformational effect of single and double 3 '-S-phosphorothiolate substitutions within deoxythymidine trinucleotides." *Magn. Reson. Chem.*, **45**: 340-345.

191. Sawai, H., Wada, M., Kouda, T. and Ozaki, A.N. (2006) "Nonenzymatic ligation of short-chained 2'-5'- or 3'-5'-linked oligoribonucleotides on 2'-5'- or 3'-5'-linked complementary templates." *ChemBioChem*, **7**: 605-611.